Master's Programme in Industrial Engineering and Management

Adapting marketing mix modelling for the retail marketing environment

A road map for development

Antti Heliste

Aalto University

MASTER'S THESIS

Aalto University
MASTER'S THESIS 2019

Adapting marketing mix modelling for the retail marketing environment

A road map for development

Antti Heliste

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Technology.
Otaniemi, 28 September 2019

Supervisor:	Timo Seppälä, D.Sc. (Tech.)
Advisors:	Hannu Laine, D.Sc. (Tech.)
	Mikko Ervasti, D.Sc. (Tech.)

Aalto University
School of Science
Master's Programme in
Industrial Engineering and Management

Aalto University
Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Antti Heliste

Title
Adapting marketing mix modelling for the retail marketing environment – A road map for development

School School of Science

Master's programme Industrial Engineering and Management

Major Operations and Service Management **Code** SCI3049

Supervisor Timo Seppälä, D.Sc. (Tech.)

Advisors Hannu Laine, D.Sc. (Tech.); Mikko Ervasti, D.Sc. (Tech.)

Level Master's thesis **Date** 28 September 2019 **Pages** 78 + 10 **Language** English

Abstract

Measuring the impact of marketing is essential for improving its performance and justifying marketing decisions to top management. However, marketers often struggle with it, even though various methods are available to them in the literature. A good starting point for marketers is the most popular method, marketing mix modelling (MMM), that is a linear regression fitted in sales and marketing data. Yet, it often suffers from various downsides, such as lack of data, deficient model forms and biases. Researchers have consequently suggested improving it through better data, better models and model validation. However, researchers mainly discuss these areas as a way to improve model accuracy rather than to widen the scope of analysis. Improving modelling granularity would enable marketers to analyse performance on lower levels and broaden their discussion on improvements. Higher granularity could particularly support the retail industry, where marketing is a complex operation because of wide product ranges, geographical reaches and customer bases. Consequently, the goal of the thesis was to analyse how the typical MMM is limited, how model developers could adjust it to meet the needs of the retail marketing environment and what impacts such adjustments would have.

We conducted the study through a combination of a literature review and a simulation. The literature review discovered that the typical MMM is limited in use in the retail environment, mainly due to its low granularity that hides the information about the structure of performance. Other flaws include, e.g., the lack of modelling in retail-specific effects, such as stock-up, and the lack of model validation. The most significant opportunity arises from increasing granularity in at least three dimensions: frequency, geography and product hierarchy. Other improvements, in turn, arise from improving accuracy through comprehensive modelling and model validation through simulation.

The simulation studied the impact of granularity on available improvement opportunities in the retail environment. A product-level model was able to reach a significant 33.1% increase in total profit compared to the unoptimised baseline. The traditional model, in turn, was only able to reach a meagre 1.7% improvement. The result supports the hypothesis that higher modelling granularity leads to more detailed and effective improvement opportunities in retail marketing. Based on the literature review and the simulation, we formed a road map for the development of MMM in the retail environment.

Keywords marketing mix modelling, retail, big data, data granularity

A! Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Antti Heliste

Työn nimi
Markkinoinnin tehokkuuden mittaamisen kehittäminen jälleenmyyntialalle – Tiekartta kehittämiseen

Korkeakoulu Perustieteiden korkeakoulu

Maisteriohjelma Tuotantotalous

Pääaine Tuotannon ja palvelujen johtaminen **Koodi** SCI3049

Valvoja Timo Seppälä, TkT

Ohjaajat Hannu Laine, TkT; Mikko Ervasti, TkT

Työn laji Diplomityö **Päiväys** 28. syyskuuta 2019 **Sivuja** 78 + 10 **Kieli** englanti

Tiivistelmä

Markkinoinnin vaikutuksen mittaaminen on välttämätöntä markkinoinnin tehostamiseksi ja markkinointipäätösten perustelemiseksi ylimmälle johdolle. Siitä huolimatta, useat yritykset epäonnistuvat siinä. Mittaamista varten on kuitenkin kehitetty menetelmiä, joista yleisin on ns. markkinointi mix mallinnus (MMM). Se kuitenkin kärsii useista ongelmista, kuten epätäydellisistä ja testaamattomista malleista. Parannusalueita on kuitenkin tunnistettu kolmella osa-alueella: parempi data, paremmat mallit ja mallin validointi. Näitä osa-alueita kuitenkin pääsääntöisesti lähestytään mallin tarkkuuden kehittämisen kannalta eikä analyysimahdollisuuksien laajentamisen kannalta. Analyysien laajuutta olisi mahdollista puolestaan parantaa paremmalla datan tarkkuudella eli granulariteetilla. Parempi granulariteetti voisi erityisesti tukea jälleenmyyntialaa, jossa markkinointi on hyvin monimuotoista ja haastavaa hallita, mm. laajan tuotevalikoiman takia. Työn tarkoituksena onkin tutkia, kuinka perinteinen markkinointi mix mallinnus on rajoittunut, kuinka sitä voidaan kehittää jälleenmyyntialalle sopivammaksi ja mitä vaikutuksia parannuksilla olisi.

Tutkimus toteutettiin kirjallisuuskatsauksen ja simulaationtutkimuksen yhdistelmänä. Kirjallisuuskatsauksessa havaittiin, että mallinnus on rajoittunut käyttömahdollisuuksissaan jälleenmyyntialalla pääsääntöisesti sen huonon granulariteetin takia, joka piilottaa informaation tehokkuuden rakenteesta. Lisäksi mm. jälleenmyyntialalla ominaisten ilmiöiden, kuten hamstraamisen, mittaaminen on puutteellista ja mallit kärsivät useista epätarkkuuksien lähteistä ja validoinnin puutteesta. Malleja olisi erityisesti mahdollista parantaa lisäämällä granulariteettia ainakin kolmessa dimensiossa: frekvenssi, alueellisuus ja tuotehierarkia. Toinen tärkeä kehitysalue on mallin tarkkuuden parantaminen mm. kattavan mallintamisen ja mallin validoinnista simuloinnilla.

Kirjallisuuskatsauksen lisäksi tehty simulaatio tutki granulariteetin vaikutuksia markkinoinnin kehitysmahdollisuuksiin jälleenmyyntiympäristössä. Tuotetason granulariteetilla onnistuttiin saavuttamaan 31.1 %:n parannus kokonaisvoitossa verrattuna lähtötilaan. Perinteisellä mallilla tulos oli vain 1.7 %. Tulos tukee hypoteesia, että parempi granulariteetti mallintamisessa tuo enemmän ja merkittävämpiä markkinointikehitysmahdollisuuksia jälleenmyyntiympäristössä. Tutkimuksen lopuksi muodostettiin tiekartta mallintamisen kehittämiseen jälleenmyyntialalle sopivammaksi.

Avainsanat markkinoinnin mittaaminen, jälleenmyynti, big data, data granulariteetti

Contents

Preface

When I began studying industrial engineering at Aalto University, I never thought I would write my master's thesis on marketing. Marketing seemed like an unanalytical and uninteresting subject that was mostly about brand and customer relationship management. My point of view was changed when I was, by chance, introduced to Juha Nuutinen, the CEO of Sellforte, and his marketing analytics company. The company approached marketing as a science and offered a marketing analytics service for retailers. I became immediately interested and eventually managed to land an internship at Sellforte.

During the internship, I learned about marketing analytics and its importance in the retail industry. I also found about the models that companies were using to measure marketing effectiveness measurement and was left wondering how they could be improved, especially to fit the needs of the retailers better. This thought and the suggestions from Juha eventually led to me writing my thesis on this subject.

Consequently, I would like to thank Juha for providing me with an opportunity to work on such an exciting and impactful topic. Moreover, I would like to thank my thesis advisors at Sellforte, Hannu Laine and Mikko Ervasti, and my thesis supervisor, Timo Seppälä, for providing me with helpful insights and feedback on my work.

Otaniemi, 28 September 2019,

Antti Heliste

Abbreviations

AMSS	Aggregate Marketing System Simulator
CEO	Chief Executive Officer
CMO	Chief Marketing Officer
Comb.	Combination
CPG	Consumer Packaged Goods
Incr.	Incremental
MMM	Marketing Mix Modelling
mROAS	Marginal Return on Advertising Spending
POS	Point Of Sales
ROAS	Return On Advertising Spending
ROI	Return On Investment
ROMI	Return On Marketing Investment
SKU	Stock Keeping Unit

1. Introduction

Throughout years, managers and scholars have called for the accountability of marketing (e.g. Hanssens and Pauwels (2016); Rust et al. (2004); Schwartz (1969); Stewart (2009)). Companies spend large sums of money in marketing each year (1.2 billion euros in Finland in 2018 alone (Rantanen and Kauppinen, 2019)) but struggle with demonstrating the financial returns. The February 2019 CMO Survey (Moorman, 2019) found out that demonstrating the impact of marketing on financial outcomes is the biggest C-suite communication challenge, 63.8% of the respondent marketers answering to facing such a problem regularly. Moreover, only 36.4% of the respondents claimed to be able to prove the impact of marketing quantitatively.

The lack of quantified information not only makes it challenging to optimise marketing spending but to rationalise it to the top management. A study conducted by Fournaise Group (2012) revealed that 80% of the 1 200 interviewed CEOs believed that marketers are too disconnected from the short-, medium-, and long-term financial realities of companies. Furthermore, an equal share admitted that they do not trust and are not very impressed by their marketers. This lack of trust and accountability affects the position of marketing in the company. A study by Homburg et al. (2015) discovered that marketing departments' influence had dropped in the past two decades over areas considered most significant for the success of the company, including pricing, new product development and strategic decisions, even though the influence of the marketing department had had the strongest positive effect on firm performance. Another study by Whitler and Morgan (2017) found out that 57% of CMOs had been in their position for less than three years.

To gain back their trust and influence, marketing departments need to start measuring the impact of marketing – as is done with any other investment. According to Stewart (2009), marketing can be identified to have three types of impacts: short-term (incremental) returns, e.g. sales, long-term (persistent) effects, e.g. customer attitudes, and real options, e.g. a higher price point enabled by a valued brand. Real options, however, are firm-specific and thus standardised measures can only be developed for short- and long-term effects. These measures must be financial as it is the language that top management understands and by which they are evaluated. One of the most commonly applied financial measures is the *return on marketing investment* (ROMI) that combines both the returns and costs of marketing into the equation

$$\text{ROMI} = \frac{\text{Incremental margin} - \text{Marketing investment}}{\text{Marketing investment}} \tag{1.1}$$

(Pauwels and Reibstein, 2017). The measure enables the comparison of efficiencies between different marketing investments. It is typically calculated by doing a baseline-lift valuation on short-term sales, where the effect of a marketing effort on sales is separated from the level of sales that would have been reached without the marketing effort. Marketers can adapt the measure to analyse different investments, for example, to measure *return on advertising spending* (ROAS). The measure is criticised by some researchers (e.g. Ambler and Roberts (2005); Rust et al. (2004)) for often only including short-term effects and disregarding any long-term effects. Thus, other measures may be needed to support it.

Despite the short-termism of ROMI, researchers and companies have done a considerable amount of research on the topic. The most commonly studied and applied method is *marketing mix modelling* (MMM), aka media mix modelling, where a regression model is fitted to historical data to represent sales as a function of advertising and marketing variables, such as media spend, number of views and product price, and control variables, such as weather, seasonality, and market competition. The outputs are typically ROAS figures for media, which can be used to adjust marketing budgets. The method is popular as it requires no experiments and is somewhat simple to implement. According to Chan and Perry (2017), a lack of granular data, however, commonly limits the scope of MMM just to studying the marketing effectiveness on a national level rather than, for example, regionally or by product category. This limitation reduces opportunities in detailed decision making and budgeting. MMM also suffers from various downsides, such as lack of data, untested assumptions of market behaviour and various biases. To counter many of these problems, Chan and Perry suggest three areas of improvement for MMM:

1. Better data
2. Better models
3. Model validation through simulation

Better data means, for example, increasing its quantity through a better granularity (Chan and Perry, 2017) whereas better models suggest using more robust models, such as Bayesian hierarchical modelling (e.g. by Jin et al. (2017), Sun et al. (2017) and Wang et al. (2017)). Model validation, in turn, means using realistic marketing system simulators (e.g. by Zhang and Vaver (2017)) to test and develop models. Chan and Perry, however, primarily discuss the impact of these improvement areas on model accuracy rather than the potential in widening the scope of analysis. Model accuracy is, of course, essential for a fact-based discussion on marketing but a model that is only capable of measuring the national-level marketing effectiveness limits the discussion to budget adjustments between marketing activities instead of allowing marketers to analyse what they could do to improve the effectiveness of each activity. A model that could measure the underlying performance structure, starting from promotions and other low-level activities, would facilitate more granular decision making and possibly have a more significant impact on marketing performance.

Increasing the granularity could especially support the retail industry. In retail, complex decision making happens on multiple levels. Retail managers not only have to adjust budgets between media, regions and product categories but also select the right products for promotion. Large retailers typically have thousands of different products, and they regularly run promotions on a selection of them. Selecting the right products and setting the optimal discounts for promotions is an intricate task as promotions can have significant side effects, for example, through stock-up, cannibalisation and traffic impacts.

As a consequence of the complexity, more than a quarter of all price promotions are estimated to result in decreased margins (Bavagnoli et al., 2015). Some researchers (e.g. Abraham and Lodish (1987), Natter et al. (2007) and Silva-Risso et al. (1999)) have attempted to develop automated promotion planners and *decision making systems* to make the promotions more efficient. However, these systems do not take into account all of the components of promotion effectiveness or do not attempt to split incremental effects between media and promotions. Overall, a gap seems to exist in research for a complete retail marketing mix model that would support the retailers in decision making on all levels with accurate marketing performance data.

1.1 Problem statement

The research problem of this master's thesis is to understand how marketing mix modelling could be improved to support the retail marketing environment and what effects such improvements would have. The objective is to provide a road map for developing retail marketing mix modelling.

Before we can improve marketing mix modelling for the retail environment, we first need to understand how it is limited in its current form. This need leads to our first research question:

> *1. What limitations do typical marketing mix models have, especially in the retail marketing environment?*

The objective of this research question is to create a summary of the main limitations of marketing mix modelling. The summary will help us find improvement opportunities for marketing mix modelling in the retail marketing environment. We can use the three areas suggested by Chan and Perry (2017) (better data, better models and model validation through simulation) but approach the problem with the primary goal of supporting the retail industry rather than just improving model accuracy. Our second research question is, consequently:

> *2. How can marketing mix modelling be improved for the retail marketing environment through:*

1. *better data,*
2. *better models, and*
3. *model validation through simulation?*

The objective of this research question is to provide a summary of improvement opportunities for marketing mix modelling in the retail marketing environment. Finally, we want to understand the impact of such improvements on retail marketing, leading to our third question:

3. How can such improvements support retail marketing management?

The objective of this question is to provide a summary of the expected impacts of improvements. The results from the three research questions can then be combined to form a complete road map highlighting the development opportunities for retail marketing mix modelling.

1.2 Thesis structure and research method

We conduct the study as a combination of a literature review and simulation study. First, we analyse the relevant literature on the retail industry, marketing effectiveness and marketing mix modelling and then identify the weaknesses and improvement opportunities of marketing mix modelling. This process allows us to answer the research questions 1 and 2. The literature review also answers the research question 3 to some extent, but, to give a complete overview of the improvement opportunities, we also do a simulation to study to their impact in practice. Finally, we combine the results from the literature review and the simulation and attempt to answer all the research questions and formulate a road map for retail marketing mix modelling. The results of the study are not only beneficial for retailers seeking to develop their marketing mix modelling but also for retailers that want to evaluate marketing mix modelling services offered by consultancies.

2. Measuring marketing effectiveness

2.1 Impact of marketing

To measure marketing, we need to understand how marketing affects consumers and companies. The purpose of marketing can be defined as the stimulation of favourable customer attitudes to increase customer demand and consequently, sales and profits (Hanssens and Pauwels, 2016). This process is depicted in Figure 2.1 by Rust et al. (2004). The marketing strategies of the company, such as media channel and product strategies, are realised through tactical actions, such as promotions and ads. These actions influence customer satisfaction, loyalty, attitudes towards the brand and other customer-centred elements, which form the *marketing assets* of the company. The marketing assets, in turn, drive the sales and market share of the company. Finally, the market impact can be measured as a financial impact and captured through various measures, such as the *return on investment* (ROI). The financial impact, in turn, affects the financial position of the company and its overall value.

From this marketing process, it is possible to identify different types of returns. Stewart (2009) suggests that marketing has three types of returns: short-term (incremental) effects, long-term (persistent) effects and real options (see Figure 2.2). Short-term effects take place soon after the marketing activity and appear as incremental sales, store visits, call centre contacts and other short-term measurements that are linked to cash flow. Long-term effects, such as customer attitudes, in turn, occur in the present but affect the market over the long term. For example, a marketing campaign might positively affect customer attitudes towards a brand immediately but only lead to sales in the far future. These accumulated marketing assets can be seen as a long-term reservoir of untapped sales.

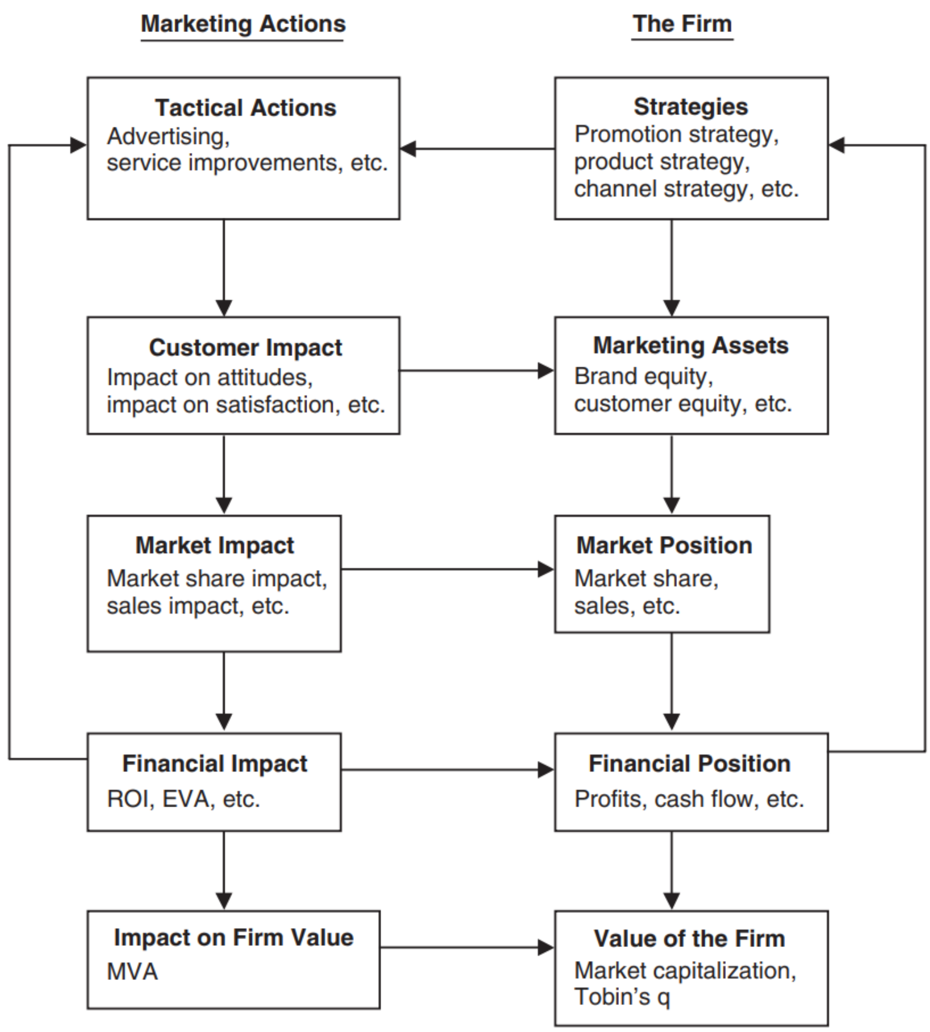


Figure 2.1. The chain of marketing productivity by Rust et al. (2004). Marketing affects the marketing assets, market and financial position and value of the company.

Finally, real options are the opportunities that marketing creates for the firm. For example, a higher-valued brand offers opportunities in price premiums or a website creates opportunities for communicating with customers and selling products. The proportion and size of the effects vary between different marketing activities. For example, Binet and Field (2013) suggest that advertising channels with broad reach (e.g. TV) are natural candidates for brand building, whereas those allowing tighter targeting (e.g. paid search) are more appropriate for short-term activation. To reach optimal results, marketers need to balance between different types of activities. Based on the data collected by the Institute of Practitioners in Advertising, Binet and Field (2013) claim that the optimal balance between the brand and activation expenditure is on average 60:40.

To optimise marketing spending and to promote the accountability of marketing, Stewart (2009) suggests developing industry-wide standardised measures for the short-term and long-term effects. The real options, however, remain specific to each company and thus are not standardisable as measures. The measures used for quantifying the short- and long-term effects should be financial. According to Stewart (2009), this is important because it is the language that top management understands and by which their performance is evaluated. Financial measures also enable comparison and benchmarking between different actions. Without such measures, making both strategic and tactical decisions becomes harder and less fact-based. For example, it would be hard to decide whether to decrease or increase the marketing budget without knowing whether the decision would increase or decrease profits. Similarly, it would be challenging to decide on the allocation of the advertising budget between different media if their effectiveness at different levels of spending is unknown. It would also be tricky to select the optimal items and their discount rates for promotion without an estimate of the incremental sales. Overall, financial measures support decision making on both the strategic and the tactical level.

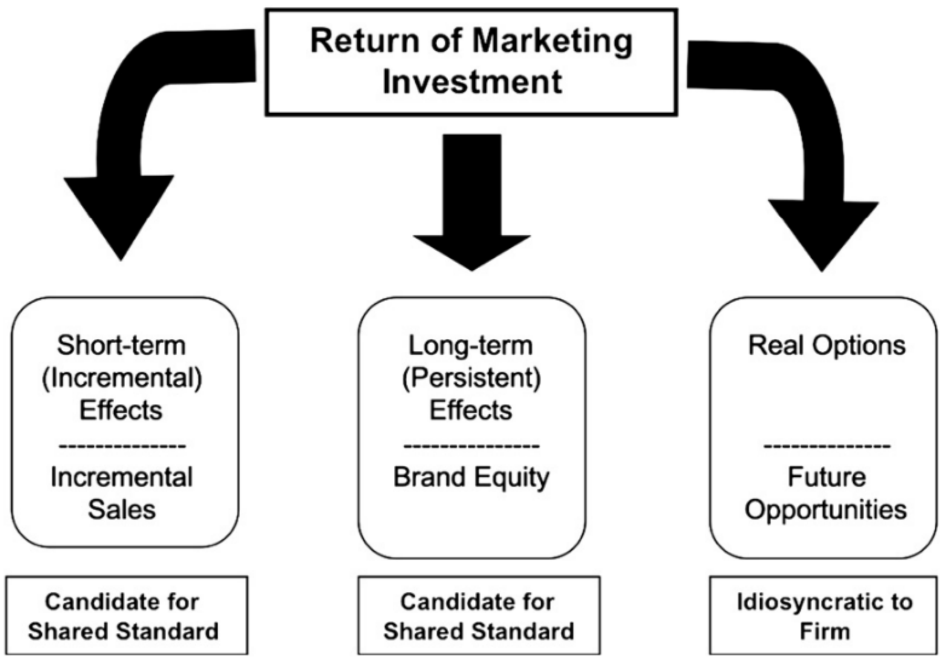


Figure 2.2. Marketing has three different outcomes: short-term (incremental) effects, long-term (persistent) effects and real options (Stewart, 2009).

2.2 Marketing return on investment

As mentioned, marketing measures should be financial. One of such measures is the widely applied *return on marketing investment* (ROMI) measure

$$\text{ROMI} = \frac{\text{Incremental margin} - \text{Marketing investment}}{\text{Marketing investment}} \tag{2.1}$$

(Pauwels and Reibstein, 2017). The ROMI compares the returns and costs of marketing, thus recognising marketing as an investment. A marketing decision with a positive ROMI can be deemed a justified investment and one with a negative ROMI unjustified. The measure also helps in comparing marketing decisions by revealing the ones with higher efficiency. The simple formula, however, leaves a lot open for interpretation. For example, how should marketers measure the incremental margins? What spending should be included in the measurement? Overall, Farris et al. (2015) identify three primary sources of variation in ROMI measurement:

The incremental returns may be measured in different ways. A typical practice is to do a *baseline–lift* valuation that splits sales into a *baseline*, which would have been reached without the specific marketing effort(s), and into an *incremental* uplift part, which can be attributed to the marketing effort(s) (Farris et al., 2015). The incremental uplift part is then used in the calculation of the ROMI. Different modellers may, however, define the baseline and incremental sales differently. A common approach (e.g. by Cain (2010)) is to define the baseline as the long-term or trend component of the sales time series that is driven by the underlying consumer preferences (i.e. the marketing assets), regular shelf price, distribution and other underlying factors, and the incremental sales as the week-to-week sales fluctuation driven by marketing and promotions.

To derive more accurate results, some models further dissect the incremental sales. For example, Silva-Risso et al. (1999) divide incremental sales of promotions into 'borrowed' and 'truly incremental' (see Figure 2.3). The borrowed sales are purchases that consumers would have done in the future but were accelerated or cannibalised by stockpiling because of the promotion. Other such side effects, such as cannibalisation and halo, may also be taken into account.

Besides incremental sales, ROMI measurement can also be based on cost-savings, conversion rates and changes in customer equity and marketing assets (Farris et al., 2015). The changes in customer equity and marketing assets can be used to estimate the long-term ROMI, although measuring the effects and setting a reliable baseline may turn out to be difficult (Stewart, 2009).

The scope of the ROMI measure can also change. For example, it can range from measuring the overall marketing impact to the impact of an individual campaign, promotion, ad or ad view. The measure can appear under new names when changing the scope, for example, as *return on advertising spending* (ROAS) when discussing advertising spending. Overall, the ROMI can be adapted to support both strategic and tactical decisions at various levels of the organisation.

The range of the ROMI can take three different forms: *total*, *incremental* and *marginal*. The total ROMI evaluates the return on all spending, the incremental ROMI the return on a specified spending increment and the marginal ROMI describes the return gained when increasing spending with one unit (Farris et al., 2015). These forms can be useful in different decisions. For example, the total ROMI can be used to evaluate whether an investment is profitable in general, whereas the incremental and marginal ROMI figures can be used to evaluate whether it is worth putting more money into an investment.

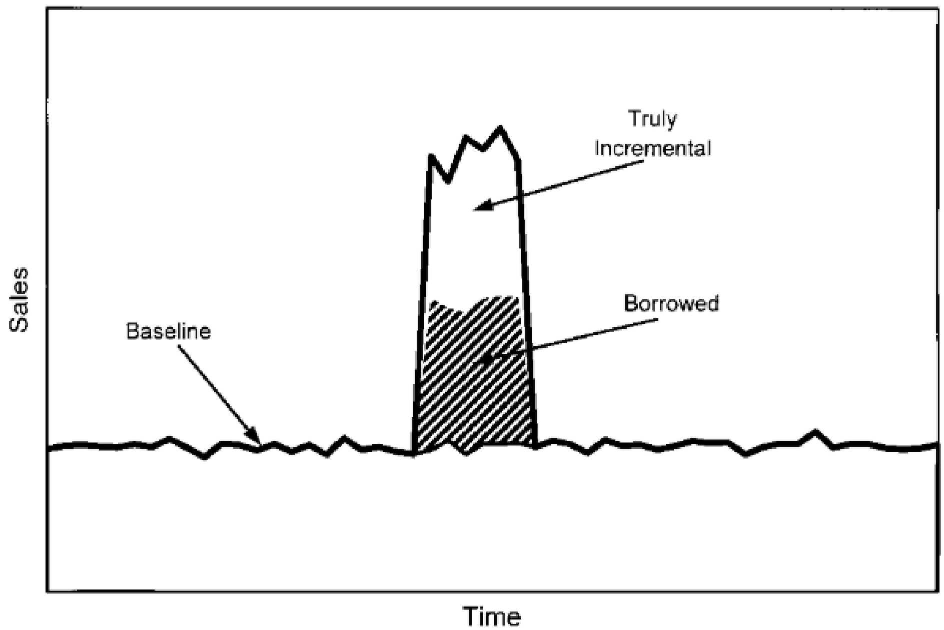
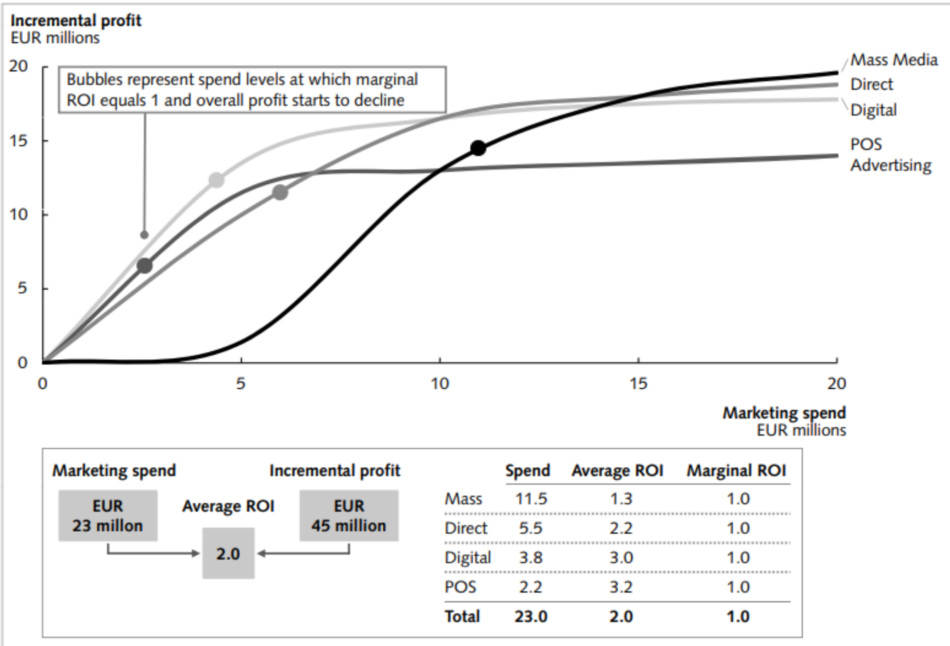


Figure 2.3. Baseline sales vs incremental sales. Incremental sales can be divided into 'borrowed' and 'truly incremental' sales. (Silva-Risso et al., 1999).

Because of all the possible variations, marketers and managers can easily misinterpret ROMI numbers. Consequently, Farris et al. (2015) recommend that marketers fully disclose their definition and measurement method of the ROMI to promote transparency and to support effective decision making.

Besides just calculating the current ROMI, marketers also often attempt to estimate the ROMI at other levels of spending. Typically, the estimate takes the form of an *S-curve* that is shaped by two inherent characteristics of media spending: the necessity to have a certain level of awareness before marketing starts to be effective and the diminishing returns of media investment (Mantia, 2015; Wang et al., 2017). An example of such curves is given in Figure 2.4 by Karmann et al. (2015). For each point on the curves, the total and marginal ROAS are calculable. The definition of ROAS is slightly different from the ROMI as the advertising spending is not subtracted from the incremental profit.

As suggested by Karmann et al. (2015), marketers can use the S-curves as a sanity check for saturation when trying to set the marketing budget. For example, if the marginal ROAS of an advertising channel drops below one, it means that for each additional euro spent on that channel less than one euro is returned. In other words, any spending after reaching the marginal ROAS of one decreases overall profits and thus is not advisable. Consequently, the curves can help marketers shift budgets towards activities with higher returns. The curves, however, may not often take into account new strategic opportunities and changes in the market and thus, should only act as a sanity check in budgeting and not as a strict guideline.



	Spend	Average ROI	Marginal ROI
Mass	11.5	1.3	1.0
Direct	5.5	2.2	1.0
Digital	3.8	3.0	1.0
POS	2.2	3.2	1.0
Total	**23.0**	**2.0**	**1.0**

Figure 2.4. An example of media S-curves (Karmann et al., 2015).

When using ROMI figures to make decisions, it is essential to distinguish effectiveness and efficiency. *Effectiveness* refers to the ability to reach a goal, whereas *efficiency* refers to the ability to do so with the least amount of resources (Hanssens and Pauwels, 2016; Pauwels and Reibstein, 2017). For companies, the goal is to make a profit. Thus, the ROMI measures the efficiency of marketing efforts rather than their effectiveness. The upper part of the ROMI equation, in turn, represents effectiveness. Although improving efficiency, in general, also leads to improving effectiveness, it may fail in some cases. Imagine two mutually exclusive marketing projects: the first one returning EUR 100 million for a EUR 50 million investment and the second one returning EUR 10 million for a EUR 2 million investment. The first one has a higher net return (€50 mill. vs €8 mill.) but a lower ROMI (100% vs 400%). Thus, the goal in marketing should be to maximise effectiveness rather than just efficiency, even if it reduces efficiency (Farris et al., 2015; Hanssens and Pauwels, 2016). One way to reach this goal is to use incremental ROMI between two choices (Farris et al., 2015). In our example, we get a 187.5% incremental ROMI ($(100-10)/(50-2) \cdot 100\% = 187.5\%$) when selecting the first project over the second one.

To reach higher profits, marketers should not blindly shift budgets toward high ROMI marketing efforts. Low ROMI operations may be saturated but effective. High ROMI operations may, in some cases, become quickly saturated, and sudden shifts towards them can lead to lower effectiveness. Moreover, the ROMI does not, by default, tell about the risks involved investments. Additional investment into an activity that has performed well in the past can always fail. Thus, marketers should consider marketing spending as an investment and carefully approach the decisions with the help relevant information, e.g. ROMI figures, S-curve estimates, forecasts and risk analyses.

Although the ROMI is a useful measure in marketing decision making, some researchers criticise it. For example, Ambler and Roberts (2005) and Rust et al. (2004) note that the ROMI often dismisses the potential impact of marketing investments on long-term marketing assets, thus yielding inaccurate results on the profitability of investments. Consequently, they recommend using other long-term measures alongside the ROMI. Despite the critique, the ROMI remains a simple and effective way to assess the short-term impact of investments. Moreover, if a company manages to quantify the financial impact of long-term effects, it can include them into the ROMI to form a complete estimate.

Overall, the ROMI can help marketers to gain back the trust of CEOs. In a study by Fournaise Group (2012), 74% of the interviewed CEOs wanted marketers to become ROI-focused. However, at least two factors hinder the adaption of the measure. Firstly, as noted by Bendle and Bagga (2016) and Whitler and Morgan (2017), marketers often misunderstand and misuse the measure, ultimately failing to speak the top management's language. Secondly, measuring the effects of marketing in practice is not simple. Over the years, researchers have studied the subject and companies have attempted to implement methods in practice. However, no simple, complete and problem-free solution is in sight. Yet, a deeper understanding of the ROMI, its usage and its measurement would certainly support managers on their path towards accountability.

2.3 Marketing mix modelling

The ROMI can help in answering many marketing questions. For example: 1) How much return do we get from spending X amount on marketing? 2) How should we allocate our marketing budget to maximise our sales? Because these questions are mainly causal, it makes sense to use commonly accepted methods for studying causality to measure the impact of marketing and the ROMI. Randomised experiments and the potential outcomes framework (Imbens and Rubin, 2015) are perhaps the most common ones of these methods and researchers regularly apply them in other fields. Randomised experiments have also been applied successfully to measure marketing effectiveness, and their cost has been brought down by the Internet and digital measurement infrastructures (Lewis and Rao, 2015). Thanks to their effectiveness, they have also become a part of the innovation process of many businesses, such as Amazon and Microsoft (Kohavi et al., 2009) and incorporated into advertising tools of Google and Facebook (Facebook, 2015; Google, 2018). However, they also have several barriers that prevent their adoption. For example, to reveal the S-curve, experiments must be done at various levels of advertising spending, which can be costly. Moreover, the need and cost of having large control groups to measure weak effects limit the opportunities in their use (Chan and Perry, 2017).

Because of the impracticality of the conventional methods for studying causality, companies often resort to using more straightforward *marketing mix modelling* (MMM), aka media mix modelling, techniques. MMM attempts to model the demand response by fitting a model, typically an OLS regression with various drivers and control factors, in historical data (Chan and Perry, 2017). The results commonly include ROAS estimates for advertising channels and a break-down of the effect of different drivers. The technique is cheap, fast and straightforward to apply as it typically involves no experimentation. Yet, the validity of the results is often questionable. The models are often based on inaccurate or unverified assumptions about the nature of marketing environment (Zhang and Vaver, 2017) and are merely capable of producing correlational, not causal results (Chan and Perry, 2017).

2.3.1 Data sources for modelling

As described by Chan and Perry (2017), MMM typically uses four types of data to fit their models:

- **response data**, which are typically volume or sales data,
- **media metrics**, which are commonly media spending data but sometimes other KPIs such as impressions and clicks,
- **marketing metrics** such as price, promotions, product distribution, and
- **control factors** such as seasonality, weather and market competition.

If suitable data are available, modellers can fit models on various levels of data aggregation, for example, ranging from the level of an individual product to the total business unit level, as presented in Figure 2.5 by Cain (2010). The lower the level of the fitting, the more detailed conclusions can be drawn from the results. However, data granularity often limits the choice. As described by Chan and Perry (2017), companies typically have granular sales data on SKU or store level but do their advertising on a higher level over an entire country. As the data types in the model should preferably be on the same aggregation level, the least granular data type often sets the granularity of the whole model. Thus, models are often fitted on the national level using weekly or monthly aggregated data. Moreover, some of the data types may be difficult to obtain. For example, companies often omit competitor data, such as marketing efforts and prices, or ad exposure data, especially for offline media, because of the difficulty of collecting reliable data. The omission can, in turn, lead to inaccuracies in modelling. Overall, the lack of (reliable) data is one of the main limitations of MMM.

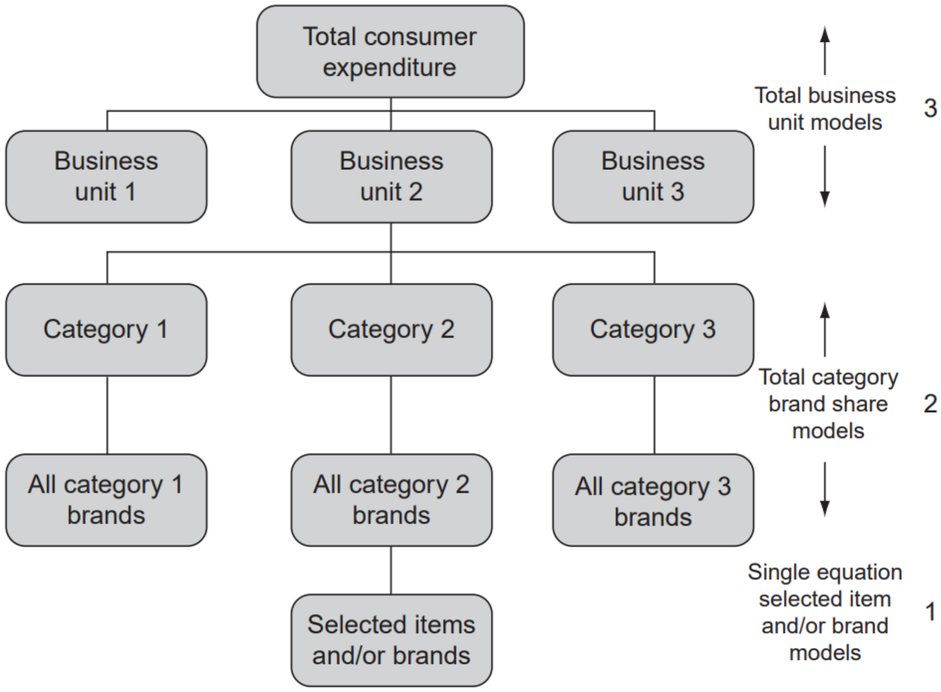


Figure 2.5. Examples of various levels on which models can be fitted (Cain, 2010).

2.3.2 Model form and transformations

Modellers (e.g. Mantia (2015)) commonly base their models on a simple additive OLS regression

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 + \cdots + \widehat{\beta}_n X_n + u, \tag{2.2}$$

where $\widehat{Y}$ is the dependent variable (usually sales volume), $X_1, X_2, \ldots, X_n$ are the explanatory variables (media metrics and control factors), u is the error term and $\widehat{\beta_0}, \widehat{\beta_1}, \ldots, \widehat{\beta_n}$ are the regression coefficients that typically represent the marginal ROAS (mROAS) of each advertising channel. The definition of ROAS is again slightly different from that of the ROMI as the advertising spending is not subtracted from the returns. The average ROAS numbers can, in turn, be derived by calculating the incremental returns at zero advertising spending and the current level.

The linear model by itself is a somewhat naive representation of the mechanics of the real market. Thus, modellers extend the model using both linear and non-linear transformations that attempt to simulate different market mechanics. The most common ones of these are the *adstock* and the *S-shape*.

Adstock describes how the awareness of an advertisement gradually decays after its ending instead of instantly disappearing. For example, a leaflet sent to households on Monday might contribute to sales later in the week as well. Mathematically, the adstock can take various forms such as the geometric adstock (Mantia, 2015; Mhitarean-Cuvsinov, 2017) or the delayed adstock (Jin et al., 2017). Figure 2.6 shows an example of the shapes of the adstock functions. The geometric model is a prevalent one and mathematically it is defined as a geometric decay function

$$\text{adstock}(x_M) = (x_1, x_1\lambda_M + x_2, (x_1\lambda_M + x_2)\lambda_M + x_3, \ldots, \sum_{j=1}^{t} x_j \lambda_M^{t-j}), \tag{2.3}$$

where x_j is the marketing metric at time j and λ_M is the media-specific decaying factor (Mantia, 2015; Mhitarean-Cuvsinov, 2017). As a measurement of 'half-time', the λ_M is also useful in selecting the optimal interval between two advertisements (Mantia, 2015).

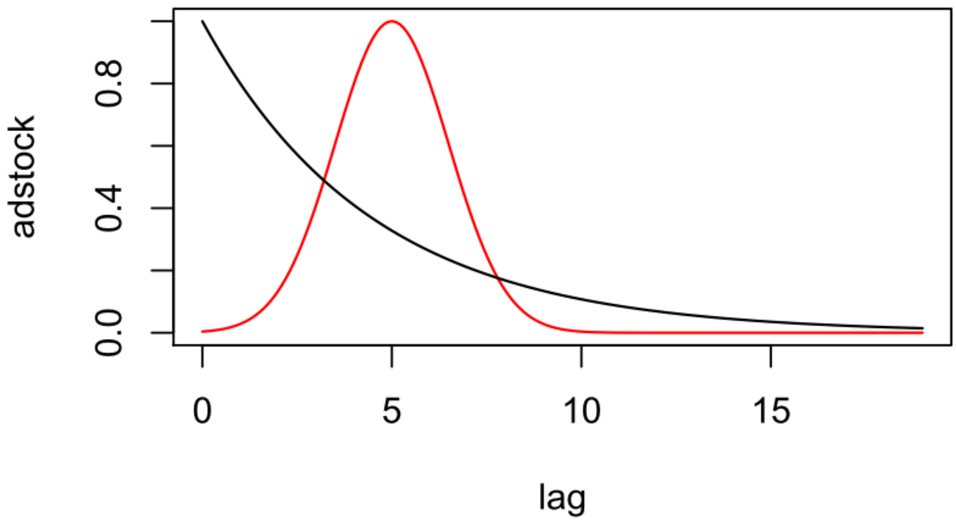


Figure 2.6. Example weight functions for geometric (black) and delayed (red) adstock (Jin et al., 2017).

S-shape represents, as described earlier, the shape of the media investment return curve. Two effects shape the curve: the diminishing returns of media investments and the necessity to have a certain level of awareness before marketing start to be effective. The S-curve helps in deciding budget allocations by revealing media channel saturation.

The S-shape can take various mathematical forms, for example, the Hill transformation

$$\text{Hill}(x_{t,m}; \mathcal{K}_m; \mathcal{S}_m) = \frac{1}{1 + (x_{t,m}/\mathcal{K}_m)^{-\mathcal{S}_m}}, \tag{2.4}$$

where $x_{t,m}$ is the value of a variable at time t for media m, $\mathcal{S}_m > 0$ is a shape parameter called as the slope and the $\mathcal{K}_m > 0$ is the half-saturation point, where $\text{Hill}(\mathcal{K}_m) = 1/2$ for any value of $\mathcal{K}_m$ or $\mathcal{S}_m$ (Jin et al., 2017). An example of the S-shape under the Hill transformation is given in Figure 2.7.

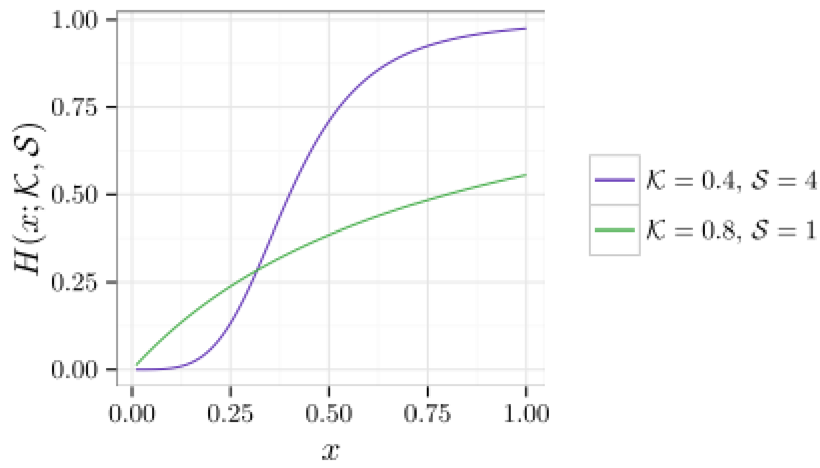


Figure 2.7. Illustration of Hill transformation with two sets of parameters (Wang et al., 2017).

The non-linear returns to scale can also be modelled through logarithmic model forms. For example, in a log-linear model, the response variable $\widehat{Y}$ is modelled as logarithmic ($\ln \widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \ldots$), whereas, in a double-logarithmic model, both the response variable and explanatory variables are modelled as logarithmic ($\ln \widehat{Y} = \widehat{\beta}_0 + \ln \widehat{\beta}_1 X_1 + \ldots$) (Mhitarean-Cuvsinov, 2017). Cain (2010), in turn, applies the transformation in practice in his work with the model form

$$\ln S_{it} = \mu_{it} + \delta_i + \sum_{j=1}^{n} \sum_{k=1}^{M} \beta_{ijk} \ln X_{kit} + \varepsilon_{it}, \tag{2.5}$$

where S_{it} is the sales of a product i over time t and X_{kit} is a marketing or economic variable k with a coefficient β_{ijk}. In turn, μ_{it} is a time-varying trend, δ_i is a seasonal index and ε_{it} is the error term. Cain's model is also, in general, rather advanced: it attempts to split sales into trend, seasonal and incremental components and takes into account various sales drivers. The result is a sales decomposition with an evolving baseline (see Figure

2.8) that can be used to analyse marketing effectiveness over time and to calculate the ROMI.

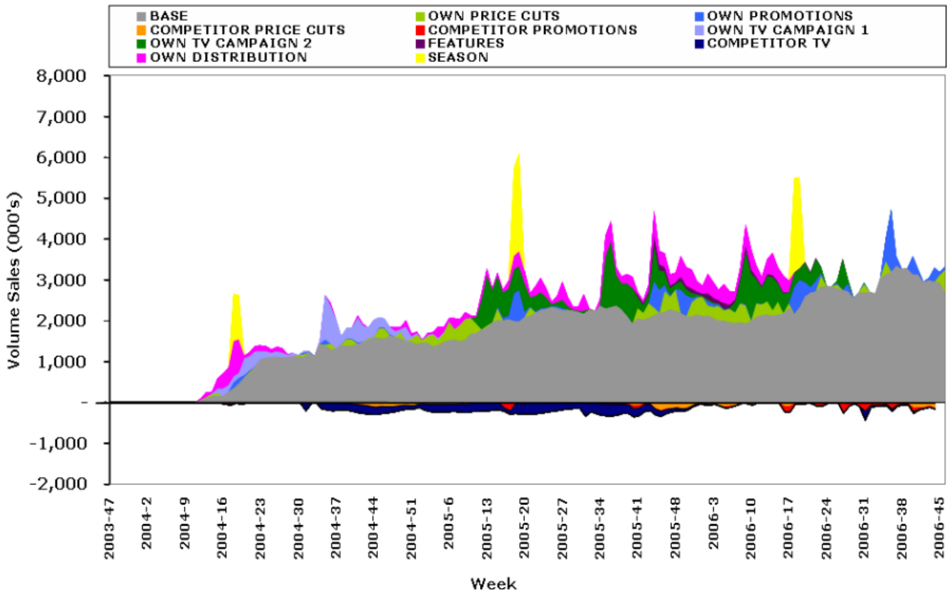


Figure 2.8. Sales decomposition with an evolving baseline for a face cleaning product (Cain, 2008).

Overall, the linear regression model is extremely flexible and modellers can use various transformations identified in the literature to adapt it to their needs.

2.3.3 Limitations of marketing mix modelling

As described by Chan and Perry (2017), the results from a regression-based marketing mix model cannot be claimed to be causal in general, but they are likely to be more trustworthy when:

- enough data are available to estimate all the parameters in the model,
- useful variability exists in the advertising levels and control variables,
- the model inputs vary independently,
- the model accounts for all the important sales drivers, and
- the model captures the causal relationship between variables.

Despite the importance of these conditions, models often fail to meet them, limiting not only the reliability of results and but also the scope of the analysis. Overall, these and many other challenges limit the opportunities in MMM:

The lack of data limits the reliability of the results and the model granularity. As mentioned in Section 2.3.1, companies typically may have accurate data on sales but lack granular data on marketing. Low granularity in one data type can limit the whole model to the same level, forcing the modeller to aggregate other data types. The aggregation leads to a data set with a low number of data points. For example, a three-year data set of weekly-aggregated data has only 156 data points. If a company has several advertising channels, each of which is modelled with several parameters, the number of data points per parameter in the model can quickly become quite low. The result might be an unstable regression fit that gives poor estimates (Chan and Perry, 2017). Many researchers have suggested that the number of observations should be at least five times the number of parameters (Leeflang et al., 2015). The lack of data can also lead to the omission of variables. For example, companies often fail to gather data about various control activities with a sales impact, such as competitor activity, leading to possible errors because of their omission (Chan and Perry, 2017).

Data reliability is also crucial for the reliability of results. If the modeller feeds inaccurate data to the model, it will only produce inaccurate results. One typical source of inaccuracies is the lack of systematic data collection. For example, Karmann et al. (2015) note that it is often a significant challenge for companies to measure their marketing spending fully. Spending may, for example, become buried in the budgets of different departments and local subsidiaries and never become registered in the main marketing department. For example, the CMO of ICA AB reported discovering significant unregistered marketing spending, amounting more than 20% of the marketing budget, as a result of a project to establish budget transparency (Bauer and Lehmann, 2015). This hidden spending also has an impact on consumers and should be taken into account models and decision making. Consequently, Karmann et al. (2015) suggest working towards full budget transparency in the company to support accurate decision making on marketing.

A limited range of data creates extrapolation uncertainty when estimating sales outside the observed budget range (Chan and Perry, 2017). For example, if an advertiser wants to estimate the average ROAS of an advertising channel, it needs to estimate the level of sales at zero spending. If no data near zero spending is available, extrapolation may produce inaccurate results if the S-curve suddenly changes outside the available range. Thus, a model may produce accurate results for the marginal ROAS but fail to produce reasonable estimates for the total ROAS.

The lack of model validation and evaluation limits model reliability. As described by Zhang and Vaver (2017), drawing causal results from MMM requires making assumptions about the nature of the marketing environment. Although these assumptions are inevitably inaccurate to some extent, they also remain unidentified or unverified in many situations. In some cases, they may also even be unverifiable. The inaccuracies and omissions of assumptions, in turn, reduce the reliability of the results. Consequently, Zhang and Vaver state that both observational and experimental methods require careful evaluation and validation. With observational methods, such as marketing mix modelling, modellers should attempt to verify the accuracy of result estimates against a source of truth, for example, from a simulation experiment.

Correlated input parameters are a typical problem with MMM: advertisers often seek to maximise the impact of marketing efforts by investing in them in a correlated way. For example, an advertising channel could be observed at a high level only when a significant discount is in place or only during a specific season. In a linear regression model, this correlation can lead to coefficient estimates with variance and thus, lead to a wrong attribution of sales to the advertising channel (Chan and Perry, 2017).

Selection bias is a major hurdle for MMM. As described by Chan and Perry (2017), it occurs when an input media variable correlates with an unobservable demand parameter that drives sales. When that parameter is left outside the regression, the model falsely attributes sales to the media instead of the unobservable parameter. The bias can be caused, for example, by:

1. Ad targeting – Especially in digital channels, selection bias occurs

when ads are shown to people who are already interested in the product, for example, when showing ads for a related search query. If the underlying interest remains unaccounted for, it can lead to false attribution.

2. Seasonality – Unknown or inaccurately modelled seasonality can lead to selection bias as well. For example, a holiday can cause an increase in sales that becomes falsely attributed to an advertising channel. Similarly, competitors' marketing efforts are not often accounted for because no data of them are often available.

3. Funnel effects – Selection bias also occurs when the level of one advertising channel affects the level of another one. For example, a TV ad might drive more search queries leading to more paid search ads.

Researchers have made some effort to counter the selection bias. For example, Chen et al. (2018) have derived a statistically principled method for correcting the bias in paid search.

Model selection and uncertainty are also significant problems with MMM. As described by Chan and Perry (2017), the modeller has to select the best form for the model from all the possible combinations of transformations and variables, e.g. whether to use geometric or delayed adstock or whether to include the daily temperature in the model. The modeller may, for example, use R^2 or predictive error to find the most accurate model. However, MMM data sets typically have a low number of data points per variable and a low signal-to-noise ration for media variables as sales tend to fluctuate much more than them. In turn, other variables, such as seasonality, price and distribution, have much stronger signals, and a model using just those variables is often able to reach a good predicting power. Since the impact of media variables can be low on the predictive power of the model, it is difficult to weed out the forms that produce inaccurate ROAS estimates. It is even possible to end up in a situation, where different models have an equally good fit but produce different estimates for the ROAS.

Short-termism is a weakness of MMM as well. Conventional MMM focuses on measuring short-term incremental sales that will consequently, only provide an estimate of the short-term ROAS. Cain (2010) repeats the critique by Ambler and Roberts (2005) and Rust et al. (2004) by stating that this focus disregards the potential brand-building and brand-eroding effects of marketing and often leads to a bias towards promotion activity, where incremental sales are more immediate than in, for example, media activity. He suggests that to measure long-term effects, models must also focus on the base sales component of the model. The base sales reflect the changes in underlying marketing assets, for example, by growing when new customers convert into loyal customers that make repeated purchases. However, according to Cain, the standard OLS regression models are only capable of imposing a fixed or deterministic baseline that prevents the analysis of the long-term impact of marketing. Consequently, he suggests the use of time series regression models to decompose the sales into short-term (incremental) and long-term base (trend) sales (see Figure 2.8).

Past performance may not be the best indicator for future performance. Conventional MMM only provides an estimate of the past performance but does not specifically attempt to estimate what the performance will be in the future. Customer behaviour and market conditions may change in the future and affect marketing performance and S-curves. Thus, the results from MMM should be used with caution when making decisions.

2.4 Retail marketing environment

The retail industry is a complicated marketing environment. Large retailers, such as grocery chains, often sell a large variety of products through an extensive network of stores. Retailers also invest heavily in advertising and the retail industry is the top advertising spender in many Western countries (Dentsu Aegis Network Ltd., 2019). The typical marketing mix modelling can support retailers in optimising this spending by quantifying marketing effectiveness. However, retail marketing has an additional level of complexity: promotions. Most retailers have wide product ranges and they frequently run discounts, special offers and other campaigns on their products. Promotions are significant value drivers for retailers.

For example, European grocery retailers made around 28% of their sales on promotion between August 2015–2016 (Eales, 2016). Promotions are also popular among customers. According to the 2016 Global Consumer Sentiment Survey by McKinsey (Magni et al., 2017), 37% of consumers around the world say they are looking for sales and promotions as a way to save money. Similarly, 30% reported waiting for products to go on sale.

Despite their importance, promotions are time-consuming and largely unoptimised. Bavagnoli et al. (2015) claim that promotions take up to 50% of category managers' working time. They also add that more than a quarter of all price promotions result in decreased margins. Such loss-making promotions not only cost money but also take up costly ad space from more effective promotions. To save managers' time and to optimise promotion mixes, retailers need to measure promotion effectiveness alongside the typical advertising effectiveness measurement. The goal of such measurement would be to answer the key questions for promotion optimisation (as identified by Bavagnoli et al. (2015)):

1. **What to promote?** – Which promotions bring the biggest profits?

2. **How to promote?** – E.g. What promotional tactics should be used to promote the product? How big should the discount be in order to maximise profit?

3. **Where to promote?** – E.g. In what media and areas should the product be promoted?

Making such optimisations requires calculating the incremental return of promotions. The measurement should include various side effects that affect the promotion effectiveness:

Stock-up Promotions can cause consumers to stockpile products or to buy them earlier than normal (Abraham and Lodish, 1987; Silva-Risso et al., 1999). The effect can appear as a dip in sales after a promotion and result in a reduced margin. For example, a lucrative promotion of a storable item can drive people to buy it in bulk and then reduce the high-margin sales later. Similarly, a promotion can cause a consumer to buy an item earlier than normal but at a lower margin. Bulk buying is fairly common: 22% of global consumers report it as one of their money-saving strategies (Magni et al., 2017).

Thus, it is important to also take into account stock-up.

Cannibalisation The promotion of certain items might reduce the sales of others (Bavagnoli et al., 2015). For example, a consumer might opt for a discounted toilet paper brand instead of their usual brand. This behaviour is likelier if the brands are close substitutes for each other. The effect can also bring down the overall profits, especially if low margin products substitute high margin products. Figure 2.9 shows an example of sales cannibalisation on a frozen potato product.

Halo As described by Bavagnoli et al. (2015), promotions affect not only the promoted item but also the traffic and baskets of customers. For example, a lucrative promotion that is advertised in the media can bring additional people to stores, increasing the total number of purchase baskets. The promotion can also increase basket sizes when people buy more products than usual because of the promotion. This effect can take place through complementarity. For example, the sales of tonics are likely to go up when gin is in promotion. This traffic and basket impact is referred to as *halo*. The strength of the halo effect varies between promotions. For example, some promotions lure in a lot of ‘cherry-pickers’ that only buy the discounted product but nothing else, reducing the average basket value. This effect can be damaging for the profitability of the promotion if it relies on the assumption that people will buy other items alongside the promoted item. On the other hand, promotions can also lure in big spenders that splurge out on other products. Overall, the impact of the halo can be significant and thus, it should be measured. Receipt data opens up a way to study the effect as it contains information about customer baskets.

Vendor funding Suppliers are interested in maximising the sales of their products. Therefore, it is not untypical for them to do trade promotions with retailers. The supplier can lower the price for the retailer and even pay the retailer a subsidy for the promotion. Supplier spending on trade promotions can be significant. According to a report by Cadent Consulting Group (2017), the marketing spending of the manufacturers of consumer packaged goods (CPG) in the USA amounted to 20% of their sales in 2016, and roughly a half of it was spent on trade promotions. Thus, the impact of the discounts and payments should be taken into account in promotion effectiveness measurement. Furthermore, suppliers also spend money on their own advertising, which could be taken into account as a control factor in modelling.

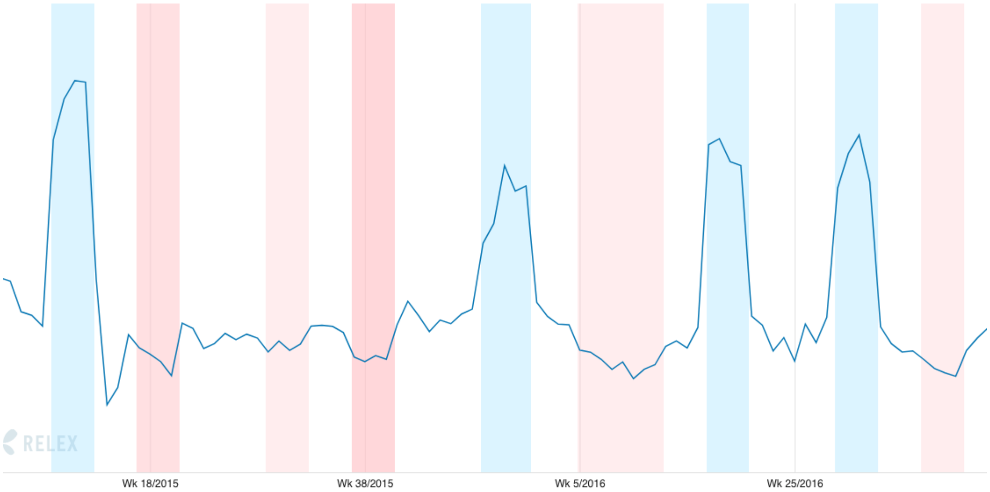


Figure 2.9. Promotions of substitutes (red periods) cannibalise the sales of a frozen potato product between 10–25% compared to surrounding weeks. The effect is, however, much weaker than the effect of promotions (blue periods), partly probably because cannibalisation is distributed across multiple products. Sales also seem to decrease after each promotion, possibly because of stock-up. (Viitanen, 2018).

When measuring promotion effectiveness, the effect of the promotion and media advertising should also be distinguished. A promotion may be run with or without media advertising. Just seeing the promotion in the store can drive consumers towards buying the promoted product and cause various sales side effects. These incremental effects should be attributed to the promotion. Media advertising, on the other hand, can bring in additional customers to stores to buy the promoted item alongside others. These sales should be attributed to the media. This definition also means that media advertising does not cannibalise sales because the additional sales would not have happened without advertising. Other side effects, however, do exist. The split between the promotion and different media allows calculating both media and promotion effectiveness and thus, making better decisions about promotions and advertising.

Overall, promotions are a complicated operation to model. Figure 2.10 highlights the various impacts that we have identified for promotions. These effects can be combined with the discount and vendor funding to calculate the exact incremental effect of promotions. Figure 2.11 shows an example break-down of effects in the form of a waterfall chart, which does not, in this case, separate the effects between the promotion and the media. In the example, a typical method would measure the uplift at only EUR 6 000 whereas a full model would measure it at EUR 9 000. Advertising

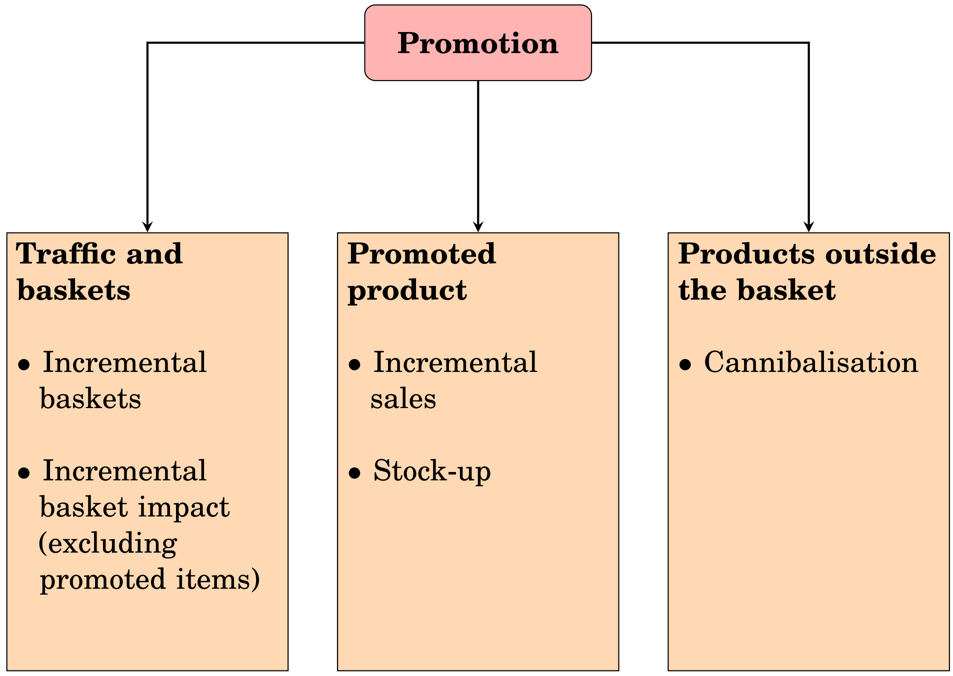


Figure 2.10. Promotion effects.

spending could be included in the waterfall as well, but calculating it for a single promoted item can be tricky, for example, when the retailer advertises several products at a time in the same media, e.g., in a leaflet. Dividing the cost of the media between the products in it does not provide a useful figure. Instead, it makes more sense to use the total incremental return to analyse promotion performance and then aggregate the numbers upwards to analyse marketing effectiveness, for example, for each leaflet.

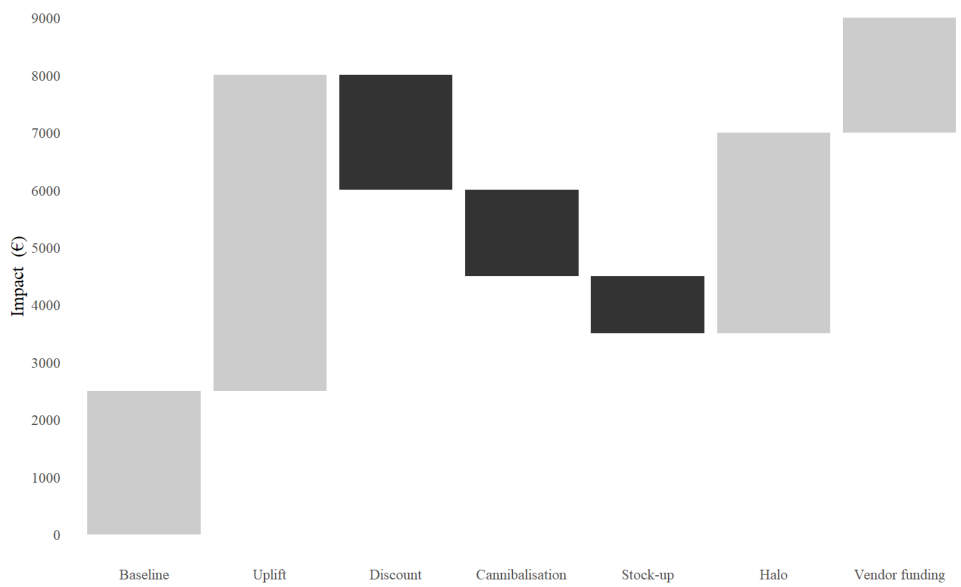


Figure 2.11. An example waterfall chart highlighting the impact of different effects on promotion effectiveness.

The accurate measurement gives retailers a more solid understanding of marketing effectiveness and helps them in picking the right products for promotion and the ways to promote and advertise them. It can also help with negotiating with suppliers on the correct level of compensation and help with estimating the optimal level of inventory needed for a promotion (Bavagnoli et al., 2015). Overall, complete effect modelling opens up opportunities in more accurate decision making. Yet, measuring all these effects and splitting the incremental sales between the promotion and different advertising channels is a difficult task. Researchers have studied promotion planning and optimisation over the years. For example, Abraham and Lodish (1987), Natter et al. (2007) and Silva-Risso et al. (1999) propose various automated promotion planners and *decision support systems* (DSS) that attempt to automate pricing, sales calendars or other promotion-related decisions. However, these systems do not take into account all of the components of promotion effectiveness or do not attempt to split uplifts between media and promotions. Consequently, lots of research remains to be done on promotion effectiveness measurement.

Another exciting improvement area comes from visualisation. Academics (e.g. Hanssens and Pauwels (2016); Leeflang et al. (2015); Pauwels (2015)) have called for the use of informative, frequently-updated dashboards for guiding marketing decisions. These dashboards are suggested to contain key marketing figures, such as the ROMI and incremental sales. A system that would combine both the top-level advertising and bottom-level promotion effectiveness measurement, visualise the essential figures on an interactive dashboard and automate decision making would undoubtedly bring value to retailers. Retailers are also good at producing at least sales data for such systems; large retailers typically have digital systems to register all purchases. Moreover, many retailers have loyalty programs that enable them to track customer purchase behaviour over time and across stores. Overall, there are notable opportunities in marketing effectiveness optimisation in retailers.

3. Improvement opportunities

As seen in the previous chapter, the traditional marketing mix modelling is a lucrative but problematic approach to studying marketing effectiveness. To improve it, Chan and Perry (2017) suggest three improvement areas:

1. Better data
2. Better models
3. Model validation through simulation

Even though these areas act a decent starting point for improvement, Chan and Perry mainly consider them from the point of view of improving model accuracy. The accuracy of results is, of course, essential for a fact-based discussion on marketing but when the model is only able to measure the national-level media effectiveness, the discussion and the available improvement opportunities will be limited as well. Thus, researchers should also pay attention to widening the scope of the discussion and finding new marketing analysis and improvement opportunities.

A simple way to widen the scope of the discussion is to improve modelling granularity. As mentioned, the traditional marketing mix modelling typically only measures the national-level marketing effectiveness, which is, of course, useful to know when deciding the budget allocation between marketing activities on a national level. However, a top-level model does not answer what contributes to marketing performance and how the effectiveness of an individual marketing activity, e.g. advertising channel, could be improved. It overlooks the importance of underlying performance factors, such as promotion mixes and geographical strategies, in the retail marketing environment. For example, an advertising channel may look ineffective only because unattractive promotions that the retailer promoted in it. The information needed for such granular analysis is lost in the *tyranny of the average* when data becomes aggregated.

As seen in Figure 3.1 by Bauer and Lehmann (2015), performance (e.g. in growth rates or ROMI numbers) may vary quite a lot between different areas, stores, product categories and other marketing units. The key to improving performance is to understand the underlying performance structure and to focus budgets on better performing and growing areas. With large retailers that advertise and promote thousands of different products in a wide range of stores located in multiple regions, the underlying performance structure is likely to have much complexity and variance and thus, more granular decision making is likely to bring in more benefit than for simpler businesses. Measuring marketing performance in such an environment, however, is not an easy task, and no complete methods exist in the literature. Consequently, the purpose of this chapter is to analyse how retailers could adapt marketing mix modelling for their unique marketing environment. We found our analysis on the three major improvement areas identified by Chan and Perry (2017).

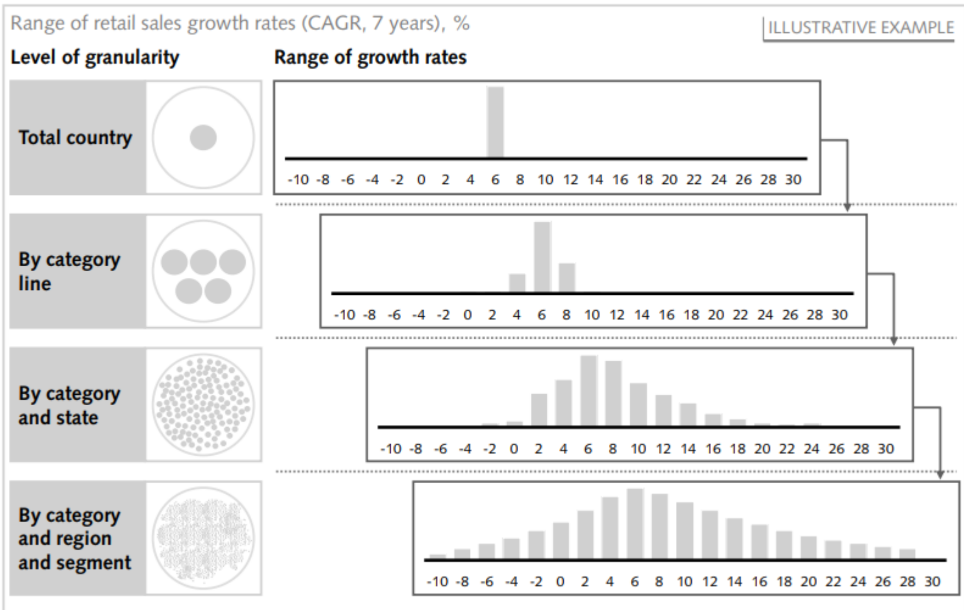


Figure 3.1. The tyranny of the average hides the underlying performance structure and improvement opportunities (Bauer and Lehmann, 2015).

3.1 Better data

As we saw in the previous chapter, many of the problems in marketing mix modelling relate to data. A typical MMM data set has a low quantity of data with low variability, which can result in an unstable fit and uncertainty in extrapolation. Similarly, the low data quality, for example in marketing spending and competitor activity, can worsen the result reliability. Besides model accuracy, the typical MMM data set also limits the model granularity to a high level because data types are not collected at equal granularity. The high-level model, in turn, hides the underlying performance structure and limits improvement opportunities, especially in retail promotion optimisation.

To improve the situation, better data are needed. Therefore, we must first consider what 'good' data implies in practice. According to Leeflang et al. (2015), good data encompasses *availability*, *quality*, *variability* and *quantity*. Availability means that data should be available. Without proper data, the model may become limited or produce inaccurate results. The quality of a data set, in turn, is measured by its *validity* and *reliability*. Validity describes whether a measure measures what it is supposed to, whereas reliability describes the accuracy of the measure. A low-quality data set will consequently produce invalid or unreliable results. Variability describes the variance of the data. In general, a higher variance increases the precision of predictor variable estimates. However, co-variation between predictors decreases precision. Quantity refers to the sheer quantity of data. A higher quantity of data will lead to more data points per parameter and, in general, to a better fit.

The simplest way to improve data quantity and variability is to increase data granularity. Data sets on lower levels are likely to have more data points and variation than the high-level aggregated data sets. Moreover, collecting highly granular data will enable deepening the level of analysis. In the retail environment, several dimensions exist for improving data quantity, for example:

Frequency As mentioned, a typical 3-year weekly-aggregated MMM data set only contains 156 data points. The low quantity of points compared to parameters reduces model reliability. Just by collecting daily rather than weekly data, retailers can gather up to seven times more data. This change should help with reaching a better fit with models. At the same time, more frequent data collection enables identifying smaller patterns in sales, for example, the optimal days to launch a promotion or to use media.

Geography Bauer and Lehmann (2015) suggest improving budgeting granularity geographically to find growth opportunities. Modelling, and consequently data collection, should be adapted to support it.

Retailers typically aggregate their MMM data sets to the national level. By moving to the area, city or store level, differences in media and promotion performance between areas become identifiable. For example, after collecting geographical budgeting data, ICA AB identified that they were overinvesting in regions with a high market share and underinvesting in regions with a low market share and consequently, changed their allocation to support growth in their low market share regions (Bauer and Lehmann, 2015). Besides granularity, geographical data collection will also increase the quantity and variability of data.

Product hierarchy Bauer and Lehmann (2015) also suggest improving data granularity in the product hierarchy. Instead of just aggregating all product data together, companies could collect detailed sales and marketing data on product categories or even on individual products. The granular data would enable the analysis of promotion effectiveness and the detailed optimisation of promotion mixes. The analysis should take into account the various promotion effects highlighted in Figure 2.11 in order to produce accurate results. Consequently, retailers also need to store receipt and vendor funding data alongside typical data types.

To increase granularity, retailers must improve data availability. Most of them already collect accurate sales and receipt data on products at checkouts. Digital marketing data are also likely to be well available thanks to automated management systems, e.g., by Google and Facebook. However, traditional marketing data may be poorly collected, for example, regarding what physical advertising was done, how much did it cost, what products were in promotion and how much did vendors pay for the promotion. The low granularity in marketing data can limit the whole model to a high level or reduce model accuracy when data are extrapolated onto a lower level. Similar problems can appear with control factors, e.g. supplier and competitor advertising, for which data are usually difficult to collect.

Improving data granularity, especially in marketing data, can, however, be difficult. For example, if advertising is done on the national level, it is difficult to dissect the budget between regions. Similarly, if an ad promotes only the brand and no products, it is impossible to assign the activity to different products despite the possible effect on their sales. Even though all marketing activity may not be assigned, the data can still be used as control factors on the lower levels to support the estimation of incremental sales on lower levels. This makes the ROMI calculable once the uplifts are aggregated up to the same level as the original marketing activity.

When improving data collection, retailers should also focus on data quality. As we saw earlier, companies often struggle with collecting accurate data of some data types, for example, competitor activity and marketing spending. If retailers feed inaccurate data into a model, it will only produce inaccurate results. Inaccurate data and results can, in turn, obscure what is happening in marketing and lead to bad decisions.

To improve the availability and quality of data, retailers should invest in data collection and storage systems and processes that support them. Data should be collected accurately and with high granularity and stored in a readily usable database format, rather than in documents. Building such systems and processes, however, can be expensive. The investments can lead to a situation where the data collection brings less profit than it costs. Thus, retailers should first focus on the data that matters the most in marketing effectiveness measurement: sales and marketing data. To measure promotion effectiveness accurately, retailers also need to gather receipt and vendor funding data. Control factor data collection can also be necessary, depending on the marketing environment. Retailers should analyse the impact of the additional variables on their business and modelling accuracy to decide whether their collection is worth it. Table 3.1 summarises the problems with data in the current system and shows the possible improvement opportunities.

3.2 Better models

Models should support granular marketing decision making in retail by providing reliable and granular results. Yet, as we saw in the previous chapter, conventional linear regression models are often limited in scope and produce unreliable results. Most of the problems arise from the lack of reliable and granular data that limits the scope of the model to a high level and makes reliable fitting difficult. Various biases and unverified assumptions also further exacerbate the problem. At the same time, existing promotion models are also deficient as they do not often split sales between media and promotions nor take into account the various side effects of promotions together. Models also often overlook long-term effects and the importance of forecasting.

Table 3.1. MMM data types, their flaws and possible improvements.

Data type	Problem	Improvement
Response data	Not many problems: POS and receipt data are usually available.	-
Media metrics	Media data often collected on low-granularity level.	Start collecting media data with higher granularity (frequency, geography, product hierarchy).
	Spending data may be hidden in budgets.	Establish budget transparency and control.
	Measurements of views and other factors can sometimes be unreliable.	Analyse the importance of each data type and what efforts are needed to improve collection.
Marketing metrics	Some data can be collected from POS data but, e.g., promotion data may be lacking.	Establish clear promotion data collection (how, where and what was promoted). Collect vendor funding data as well.
Control factors	Measuring some factors, especially market competition, may be difficult in practice.	Analyse the importance of each data type and what efforts are needed to improve collection.

Although companies could fix some of the problems in modelling with better data, model development is needed as well to improve modelling accuracy to support both retail advertising and promotion decision making. This need is why we in this section first analyse the requirements for models in the retail environment and then suggest improvements.

3.2.1 Scope of retail decision making

In the retail marketing environment, marketing mix modelling must support two types of decision making: top-level marketing mix management and bottom-level promotion management.

On the top level, marketing budget sizing and allocation are major strategic questions. They are large in scale and have a significant impact on the profitability of the company. As described earlier, marketing mix modelling can support this decision making with the S-curve analysis that acts as a sanity check for setting the spending levels. However, the typical S-curve analysis is often only limited to measuring the national-level marketing effectiveness and overlooks opportunities in, for example, regional and category-level optimisations. Some regions or product categories may be growing faster than others and require more investment. Similarly, some media may work better in one region than in others. Geographical or category models could support such decision making. Various model forms for such analyses exist in the literature, for example, by Cain (2010), Sun et al. (2017) and Wang et al. (2017). Models can, of course, be adjusted based on which granularity decision making needs to take place.

Besides marketing budget allocations, retailers need to manage their promotion complexity (i.e. what to promote? how to promote? where to promote?). Answering the three main questions of promotion management requires measuring the incremental returns of promotions (including all side effects) and then finding the optimal products, timings, discount levels and advertising methods. To calculate the incremental returns, companies need to develop granular promotion models. Promotion models have been studied and developed in the literature, for example, by Abraham and Lodish (1987), Cain (2010), Natter et al. (2007) and Silva-Risso et al. (1999). However, their models do not take into account the various side effects of promotions highlighted in Figure 2.11 or attempt to split the uplifts between media and promotions. Thus, more complete models are needed. Promotion models should be able to produce a waterfall breakdown of promotions to calculate their exact incremental effect. Moreover, they should be able to distinguish the impact of the media from the promotion itself.

Overall, marketing mix modelling should support decision making in both marketing and promotion mix management. Table 3.2 summarises the critical tasks in both and the needed information for decision making. To produce information for both tasks, retailers need advanced models. One option is to develop different models for different tasks. Another one is to create a bottom-level promotion model and then aggregate the results upwards to yield top-level results. This *bottom-up ROMI* approach preserves the information about the underlying performance structure, enabling managers to understand what contributes to performance and to make highly granular decisions.

Table 3.2. Tasks and requirements of retail marketing.

Area	Tasks	Needed information
Marketing mix management	Marketing budget allocation (channel, area & category)	S-curve, ROMI
Promotion mix management	Promotion mix optimisation (what, where & when?)	Incremental returns (waterfall)

3.2.2 Model form improvements

To fulfil the requirements and to improve accuracy, the model form also has to change. One approach is to try adapting the typical linear regression model to account for all the effects and to counter all the biases. However, other model forms can open up more opportunities. Chan and Perry (2017) suggest using Bayesian models instead of the typical linear regression to gain several benefits, such as:

- Ability to use informative priors for parameters
- Ability to handle complex models
- Ability to report on both parameter and model uncertainty
- Ability to propagate uncertainty to optimisation statements

Bayesian modelling takes a somewhat different approach than linear regression. Bayesian inference treats model parameters as random variables and rather than producing a single result, it gives out a posterior distribution of the parameters based on the data and the prior distribution of parameters (Jin et al., 2017). This distribution gives the managers an understanding of the uncertainty of results, allowing them to analyse how much they can trust the results and what is the impact of possible errors. Managers can also derive single number results from the posterior distribution if they need to, for example, by calculating the mean, median or mode of the posterior distribution (Jin et al., 2017). Figure 3.2 shows an example of posterior distributions of ROAS and mROAS.

The selection of prior distributions can have a significant impact on the posterior distribution. Prior distributions about parameters can be either *informative* or *uninformative*. Informative distributions are specific and based on past experiments, studies, results or information, whereas

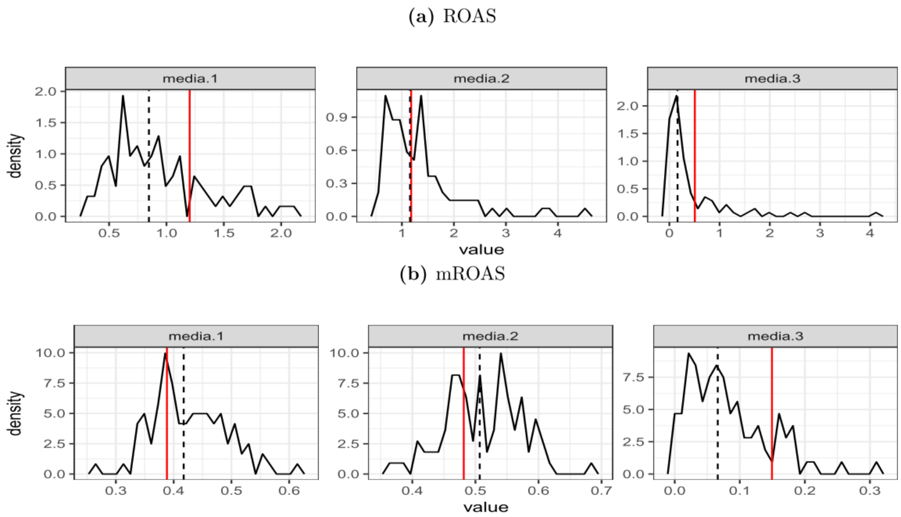


Figure 3.2. Example posterior distributions of ROAS and mROAS. The black curve is the posterior density, the dashed line is the posterior median and the red line is the truth (Jin et al., 2017).

uninformative priors are vague or general. Informative priors can improve accuracy in marketing mix modelling, especially when the data set is limited, and the number of data points per parameter is low (Chan and Perry, 2017). However, the granularity of the data set also plays an important role. For example, models by Sun et al. (2017) and Wang et al. (2017) take advantage of geographical and categorical data, respectively, to gain more credible intervals for the parameters than with a national-level data set. Such granular models also support the granular decision making needed in the retail marketing environment.

Bayesian modelling could be taken beyond geographical and categorical models to the promotional level. This move would enable using the bottom-up ROMI approach with the benefits of Bayesian modelling. However, the use of Bayesian modelling in retail promotion seems rather unstudied, although some related studies exist. For example, Becker et al. (2014) use a machine learning Bayesian approach with linear regression to measure promotion impacts in a campaign by a travel and hospitality chain. The method provides several opportunities, such as campaign estimation and accurate ROI calculation without experiments. In general, however, Bayesian modelling in promotion modelling needs to be studied more. Studies on Bayesian media modelling by Jin et al. (2017), Sun et al. (2017) and Wang et al. (2017), and promotion decision support systems by Abraham and Lodish (1987), Natter et al. (2007) and Silva-Risso et al. (1999) most likely act as a solid starting point for development.

When adapting Bayesian modelling for promotions, the various side effects of promotions should be measured as well. Luckily, some research exists on measuring these effects. For example, Silva-Risso et al. (1999) measure stock-up in their promotion model to calculate true sales uplifts and Herrala (2018) introduces a machine learning method for measuring cannibalisation between products. Basket impacts have also been studied by, for example, McAlister et al. (2009) who attempt to identify cherry-picked brands and Hruschka et al. (1999) and Manchanda et al. (1999) who attempt to measure cross-category promotion effects. Finally, Cain (2010) represents a marketing mix model that is able to split sales between baseline, promotion, media and other effects. Despite all this research, a model that would integrate all these effects is missing. Thus, research must be done in this area as well.

Model developers also need to pay attention to fixing biases. For example, selection bias can lead to false attribution of sales, through ad targeting, unknown or inaccurately modelled factors and funnel effects. In marketing mix modelling, research has been done by Chen et al. (2018) on correcting the selection bias on paid search. Similar adjustments may have to be needed depending on the marketing environment.

Modellers often disregard the long-term effect measurement in MMM. Consequently, Cain (2010) suggests using time series regression models that decompose sales into short-term (incremental) and long-term (base) sales. From this the base sales, the model should separate the effect of underlying factors that are not related to the retailer's marketing, such as distribution changes, seasonalities and competitor activity, to isolate the change in customer behaviour. The remaining baseline, in turn, reflects how new customers turn to loyal customers and how loyal customers stop shopping at the retailer, giving us an estimate of the long-term impact of marketing. Cain also describes a model that can provide such a decomposition, although it does not contribute changes in base sales to any specific media activity. Optimally, a model would allocate the changes to different marketing activities, enabling the calculation of a complete ROMI estimate.

Similarly to base sales, loyalty program data can help with identifying long-term patterns. The longitudinal data also enables the clustering of customers and creating of segment-specific marketing strategies (Allaway et al., 2006). However, to ensure accurate data collection, loyalty data should preferably be collected every time a loyalty program customer makes a purchase. If the retailer uses loyalty cards to measure the purchases, users must be kept card-loyal. Mauri (2003) suggests that customers can be kept card-loyal by providing them with promotions, discounts, points and other promotional rewards that are exclusively targeted at them. The size of the reward does not seem to matter but rather whether they exist at all and whether they are targeted. Overall, various approaches exist in measuring the long-term effects of marketing.

The final improvement area is forecasting. Most models focus on measuring past performance, which, however, might not always reflect future performance. For example, promotions may perform differently at different times, S-curves can change their form, and underlying customer preferences may change over time. Forecasting in the retail marketing environment can, thus, provide many benefits. On the product level, it can help with estimating the correct level of stock needed to minimise the chance of running out of stock and overstocking (Bavagnoli et al., 2015; Cooper et al., 1999). With new products with no historical data, it can also with planning the optimal marketing mix for the launch (Luan and Sudhir, 2010). Forecasts of promotions and promotions strategies, coupled with factual information about past performance, can also help retailers to convince suppliers to change their funding strategies (Bavagnoli et al., 2015). For example, retailers may ask suppliers to shift their funding from products that have or are likely to be margin-negative toward more promising items. Similarly, forecasts can help retailers to negotiate the correct amount of funding for a promotion. Finally, forecasting can also help with analysing the potential impact of changes in total sales. For example, Poh and Jašić (1998) apply neural networks to model the complex underlying product interrelations and effects and to forecast total sales. They also apply sensitivity analysis to analyse the impact of different marketing factors on it. Such a forecast that takes into account the interrelations could also help with promotion mix optimisation. In summary, forecasting is an area that should be more closely incorporated into marketing mix modelling.

Overall, several improvement opportunities exist in adapting marketing mix models for the retail environment. Table 3.3 lists the identified problems and improvement opportunities.

Table 3.3. Models and their improvement opportunities.

Problem	Improvement
Typical linear regression is limited	Bayesian models allow taking advantage of past information and provide useful confidence distributions for results.
Limited scope of the model	Improve model granularity, preferably down to the promotion level. Aggregate incremental effects upwards.
Limited promotion measurement	Develop a method to measure all the side effects of promotion (cannibalisation, halo, stock-up) and to split uplifts between the promotion and media.
Various biases	Bias corrections in the model form depending on the relevant biases.
Lack of long-term measurement	Attempt to capture the long-term effects of marketing through the baseline and loyalty data.
Lack of forecasting	Develop forecasting methods that can estimate future promotion and marketing performance.

3.3 Model validation through simulation

To support accurate decision making, model developers need to validate their models properly. Typical marketing mix models typically involve numerous unverified assumptions about the nature of the marketing environment and are vulnerable to various biases. Just getting a proper fit is no guarantee for model accuracy as different models forms may fit equally well in the data, especially if the signal-to-noise ratio is low (Chan and Perry, 2017). Thus, proper testing is needed. Zhang and Vaver (2017) state that comparing the model results to a source of truth is required for verifying result accuracy, especially in observational methods such as MMM. A cheap and flexible way is to do this is to simulate realistic data sets where the real marketing effectiveness is known and then feed them to the models for comparison. This way, model developers can iteratively test their models under different assumptions of the market behaviour, for example, with various level of selection bias, to improve them. Proper testing and understanding of the capabilities of the model will, in turn, give managers more confidence in the model. However, the major downside of simulation is that it is most likely impossible for a simulation to imitate the real marketing environment fully and consequently, good performance with a simulator does not automatically guarantee performance in a real marketing environment. However, by combining both simulations and traditional model verification methods, model developers are likelier to reach better results than with a mere fit analysis.

Zhang and Vaver (2017) from Google have done notable work on developing a simulator for testing marketing mix models. Their Aggregate Marketing System Simulator (AMSS) is an openly available tool that simulates realistic marketing system data sets that can be used to test MMM accuracy. The software has various adjustable marketing variables (e.g. media channels, population sizes, media effectiveness), making it extremely flexible. The simulator produces data sets that include marketing behaviour data (e.g. ad spend, website clicks, views) and customer behaviour data (e.g. purchases, customer sentiment). The tool can also accurately estimate the real ROAS of each advertising channel in the data set, making it possible to compare an estimate from a model to the real one. Overall, the controllability of the various factors and measurability of the ROAS makes the tool useful for experimenting with marketing mix models. Thus, it provides value for both the developers of MMM and companies interested in buying MMM services for consultancies. A more detailed description of the simulator can be found in Appendix A.1.

4. Research method

In the previous chapter, we identified several areas for improvement in marketing mix modelling: improving granularity, comprehensive effect modelling, improving data quality, correcting biases, forecasting and model validation. Now, we want to study how these improvements can support retail advertising and promotions management. Based on the literature, we can conclude that improving data and models with proper validation will increase model accuracy. The benefits of modelling granularity in marketing development, however, are somewhat unstudied and a key interest for the retail industry where the complexity and variance in performance are likely to increase radically with granularity.

To understand the benefits of granularity, we conduct an experiment to study what marketing improvement opportunities better granularity opens up and what is the impact of these improvement opportunities. Our hypothesis is that higher modelling granularity will open up more detailed and effective marketing improvement opportunities in the retail marketing environment. We conduct the experiment through a simulation where we generate a data set of a retail marketing system and then attempt to improve the marketing performance and overall profit by making improvements at various levels of granularity. We find the improvement opportunities by first simulating counterfactuals for the original data set to calculate the underlying incremental effects of the media and promotions. Then, we aggregate the data to match the granularity of three previously discussed models: *basic marketing mix model* (weekly national-level data), *geo-category model* (daily regional- and category-level data) and *product model* (daily product-level data). This aggregation gives us the data sets that would be available to a marketer applying an accurate MMM at different levels of granularity. After the aggregation, we attempt to find improvement opportunities at each level only using the available information. We then implement the improvements and rerun the simulation to see how much the performance improved.

The simulation provides many advantages compared to real-life experimentation and modelling. With simulation, we can calculate accurate estimates of the media and promotion performance that even include long-term effects. In a real-life experiment, in turn, the true underlying performance always remains unknown. Simulation also helps us in eliminating the impact of the model selection and accuracy by giving us equally accurate information about the underlying incremental effects on all aggregation levels. In other words, we can at each level simulate the situation that a marketer would be in if their marking mix model was accurate. Thus, we can isolate the impact of granularity on performance. Simulation is also a cost-effective, flexible and reproducible way to conduct this study.

If the hypothesis is proven, it will help us in describing the benefits of granularity for retailers and drawing the road map for development. The experiment also demonstrates what improvement opportunities the three different models can help retailers to identify and helps us to describe their use better. The following two sections describe the simulated marketing environment and the experiment in more detail but the basic outline of the experiment is as follows:

1. We create the initial data set by simulating marketing and sales data with the Google AMSS for 50 products, which are divided into 10 categories and sold in 2 regions, over 2.5 years. Each product has its own unique sales and marketing behaviour. In total, the process yields us 100 different data sets.

2. We generate the required counterfactuals to calculate the base sales and incremental effects for promotion and media for each product.

3. We drop the burn-in period of six months from the data sets and then aggregate the data to match the levels of the three selected models. We then find improvement opportunities in the marketing environment at each level of aggregation using the first 12 months of the time series.

4. We generate data sets for each model with the aggregation-specific improvements implemented during the final year.

5. We compare the profits and ROAS numbers from the final year to those of the original data set.

4.1 Marketing environment

For the experiment, we simulate a realistic retail marketing environment. We simulate the environment with the Aggregate Marketing System Simulator by Zhang and Vaver (2017) that imitates realistic customer behaviour and marketing effects, such as the S-curve, and allows us flexibly to adjust various marketing factors.

Our simulation has a retailer that has both online and physical presence. The retailer has a product range of 50 different products, each of which belongs to one of the ten product categories. The product range is rather small compared to that of a large retailer, but it represents a portion of it or the product range of small scale retailer. Each product in the product range has a distinct purchasing pattern, marketing effectiveness and margin. The retailer sells the products in two different regions. The first of these regions has a population of 5.5 million people, while the second one has a population of 10 million. Market behaviour and consumer preferences are different between regions.

The retailer regularly advertises the products in two media: TV and paid search. The advertising budget is split quite equally between the two media and the two regions. TV advertising consists of single product ads, and it has a strong brand-building long-term effect but also a weak short-term sales-driving force. Its cost varies with region and week. TV is also implemented using the traditional media module of the AMSS, meaning that it could represent almost any product-level advertising channel.

Paid search, in turn, is an imitation of the Google paid search functionality, where advertisers have to bid against each other. Contrary to TV, paid search strongly drives sales as it guides consumers to the retailer's website but lacks a brand-building effect. For simplicity, we make the retailer win all bidding auctions but force it to pay the maximum bid, which varies with product, region and season.

We selected these two media because they represent two very different types of media and because they can be applied for the promotion of individual products. For example, a TV ad can be done on one single product whereas a leaflet typically contains several products and thus, the cost is split between them. The product-specific advertising approach also allows us do to more granular adjustments than, for example, with category-wide advertising.

Besides advertising, we set the retailer to run discount promotions alongside advertising. These range from 10% to 40%. For simplicity, we also treat the product price directly as the product margin. Most promotions are set to last two weeks, and they are always supported by media activity. During the promotion weeks, TV ads are run on Mondays, Fridays and Saturdays and paid search is used daily. The promotion periods are organised with category-specific patterns quite evenly over the year without paying attention to product seasonalities. This approach gives us an environment that is largely unoptimised.

We simulate daily data for each product in each region for 2.5 years, amounting to 100 data sets, each with 910 rows of data. The first six months are used as a burn-in period during which the market behaviour stabilises and are discarded. The first complete year after the burn-in period, in turn, is used for finding the improvement opportunities and the second year as a baseline for the improvement opportunities.

For simplicity, we also generate each product-region data set independently. The products have no interdependencies and thus have no cannibalisation and halo. The effect of stock-up, however, is captured by the counterfactual generation.

Figure 4.1 highlights the structure of the marketing environment.

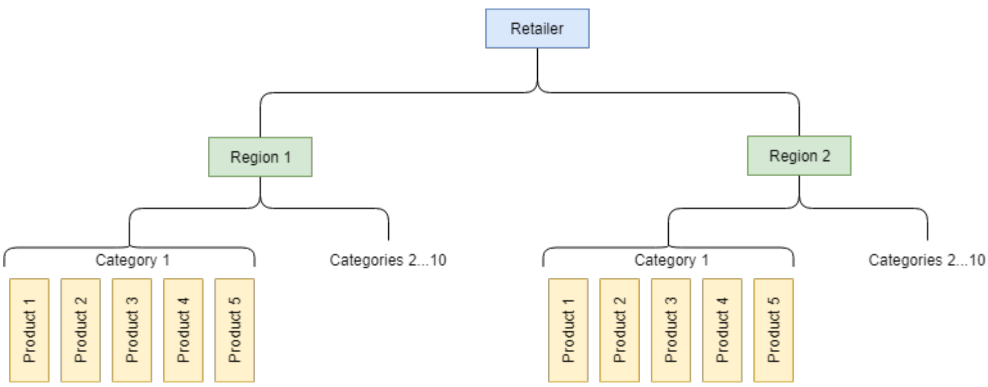


Figure 4.1. The structure of the retailer marketing environment.

4.2 Simulation

The AMSS generates a data set for each product containing various marketing variables over time. For our experiment, the most important of these are the revenue and media spending because they can be used to calculate the media ROAS figures. To calculate the true effects of marketing, we need to simulate the same marketing environment again but with adjusted factors.

We begin the experiment by first selecting the needed variables and then using them to generate the base data set from which we attempt to find improvement opportunities. The data set consists of various marketing variables over time but does not include the ROAS and incremental return needed for finding improvement opportunities. For that purpose, we need to do a baseline–uplift split by generating counterfactuals of the same data set. Overall, we need to do the following runs:

1. No marketing spending or promotion activity
2. No marketing spending
3. No search spending

These runs can be used to calculate the total decomposition of the total margin:

$$\text{Total margin} = \text{Baseline} + \text{Promo} + \text{TV} + \text{Search} \tag{4.1}$$

The first run gives us the baseline whereas the difference between the second and first one gives the effect of the promotion. The difference between the second and third run, in turn, gives us the effect of TV. Finally, the difference between the original data set and the third run gives us the effect of paid search. The incremental returns can then be used to calculate the ROAS numbers for advertising spending. The disaggregation is a reasonably accurate approximation of the different effects, although some random elements of the simulator and the attribution of the weak joint effects to paid search introduce small inaccuracies to the estimates. To increase accuracy, we could simulate counterfactuals several times but, based on our experience, one simulation is precise enough to identify ROAS numbers within the accuracy of one decimal place. Such a level of accuracy is adequate in our experiment to find improvement opportunities.

Once we have estimated the incremental effects in the data set, we can aggregate the data to match the aggregation levels of the three identified models: weekly–national, daily–region–category and daily–product. This aggregation produces the data sets that would be available to a marketer applying an accurate MMM at different levels of aggregation. After the aggregation, we can use the ROAS numbers and incremental uplifts from the first 12 months after the burn-in period to find improvement opportunities at the level of aggregation, for example, through better budget allocation and timing. We then resimulate the base data set but with the improvements implemented during the last 12 months. To find the impact of the improvements, we also need to calculate the counterfactuals for the new data sets. We can then compare the performance from the last 12 months of the new data sets to that from the corresponding range of the base set to see what improvements were made in terms of ROMI and total profit.

As each data set and counterfactual consists of 100 product data sets, we will have to simulate a total of 1 600 product data sets during the experiment. The process is quite slow by default but we can speed it up significantly through multi-threading. The code and settings used in the simulation are available on BitBucket at `https://bitbucket.org/heliste/master-thesis-antti-heliste`. The simulated data files, however, were too large to be publicly shared.

5. Results

5.1 Time series decomposition and analysis

Now that we have generated the entire data set with all the required counterfactuals, we can start analysing it. First, we slice away the burn-in period of half a year, leaving us with a full one-year data set that begins from the start of July. Then, we analyse the overall sales and the patterns in it. Figure 5.1 shows the national and regional weekly total margin decomposed into the baseline, promotion, TV and search effects alongside media spending. The national time series represents a decomposition that a well-developed but a national-level marketing mix model could produce. The regional decompositions, in turn, represent figures that a geo-category model could have produced.

In the decompositions, the *baseline* shows the level of margin that would have been reached without any promotions or advertising. The baseline reflects the underlying seasonalities and consumer preferences. For example, we can see that Christmas and May are high seasons whereas the beginning of the year and July are low seasons. The *promo* line, in turn, shows the level of margin that would have been reached with promotions but without any media activity. Overall, promotions seem only to have a slight effect on total margin. The *tv* and *search* lines, in turn, show the estimated effect of TV and paid search. The effect of paid search appears to be slightly stronger than that of TV. Media, in general, seems to have a large positive impact on the total margin. The effect, however, varies over time, partly because the number of promoted products and media spending keep changing. On the other hand, it varies with the marketing efficiency of the underlying products. Unattractive products are less efficient, and they can also attract less paid clicks, reducing paid search spending.

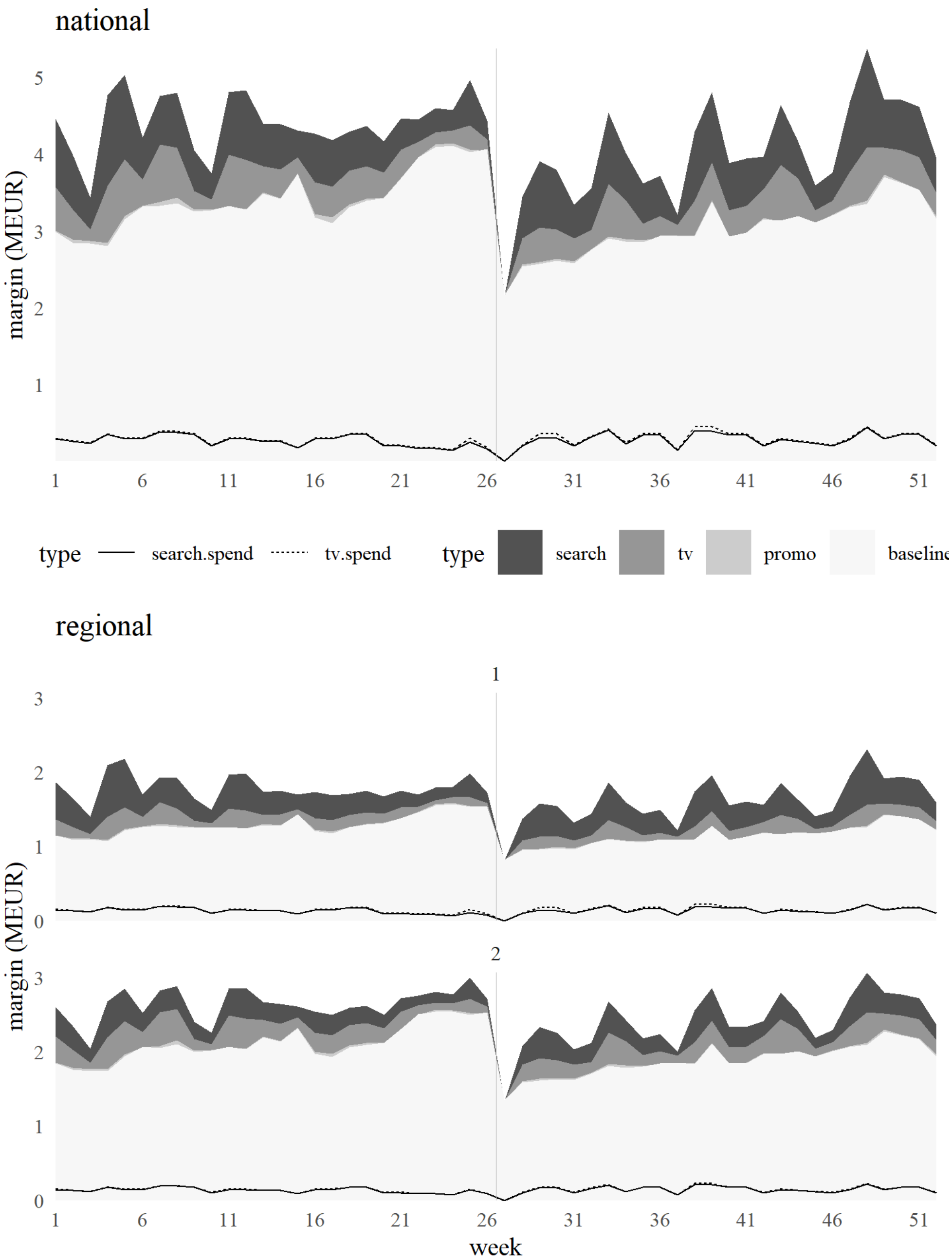


Figure 5.1. Weekly national and regional baseline-promotion-media disaggregation with media spending. The grey line indicates a start of a new year.

We can try to explain the variation in the size of the media effect at different time points by plotting the weekly ROAS over time. Figure 5.2 shows us the weekly total, paid search and TV ROAS figures. We can see that, during Christmas, TV and paid search were not incredibly efficient, probably because of higher media costs and the products promoted in them. Christmas was also partly weak because of the slightly lower number of promotions than usually. We can also see that the high sales peaks have excellent efficiency in all media. We can also see that the total ROAS stays around two over year, meaning that the overall return from marketing is positive. Although the high-level model allowed us to see media efficiency over time, it does not tell us how categories or products performed at each time point and reduces our improvement opportunities.

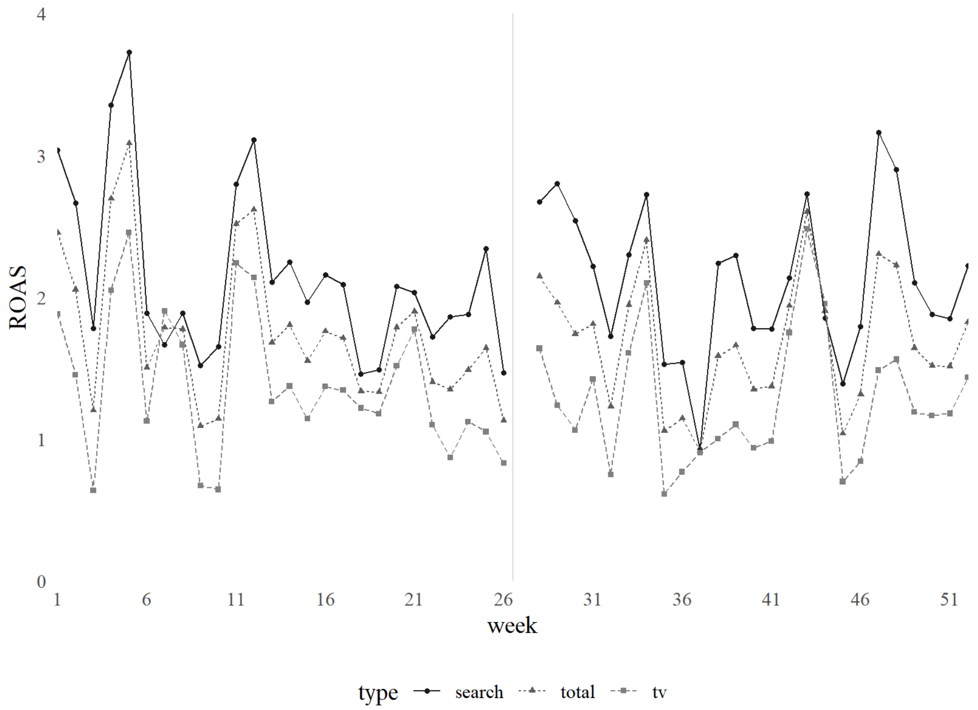


Figure 5.2. Weekly ROAS over time for paid search, TV and all advertising.

With a geo-category model, we can analyse the regional patterns. In Figure 5.1, we can see that Region 2 has higher sales, likely due to the larger population. We can also see that TV is stronger in Region 2 than in Region 1 and search is stronger in Region 1 than in Region 2. Seasonal patterns on the total level, however, are quite similar except Christmas has a stronger baseline in Region 2.

Our selected geo-category model also takes advantage of daily rather than weekly data, allowing us to analyse intraweek patterns. Figure 5.3 shows us that Wednesday, Friday and Saturday have higher baseline sales in general. This information suggests that we should focus our media on those days of the week. Without daily granularity, we could not have analysed this.

The geo-category model also enables us to analyse product categories. Figure 5.4 shows the daily category-level sparklines. We can see that category sales surge with media and promotion activity. We can also explain why media was less efficient during Christmas: some of the best categories (5 and 9) were not in promotion while some of the promoted categories (6 and 10) had a low season during it. With better promotion selection and timing, the results would have been better. We can also spot

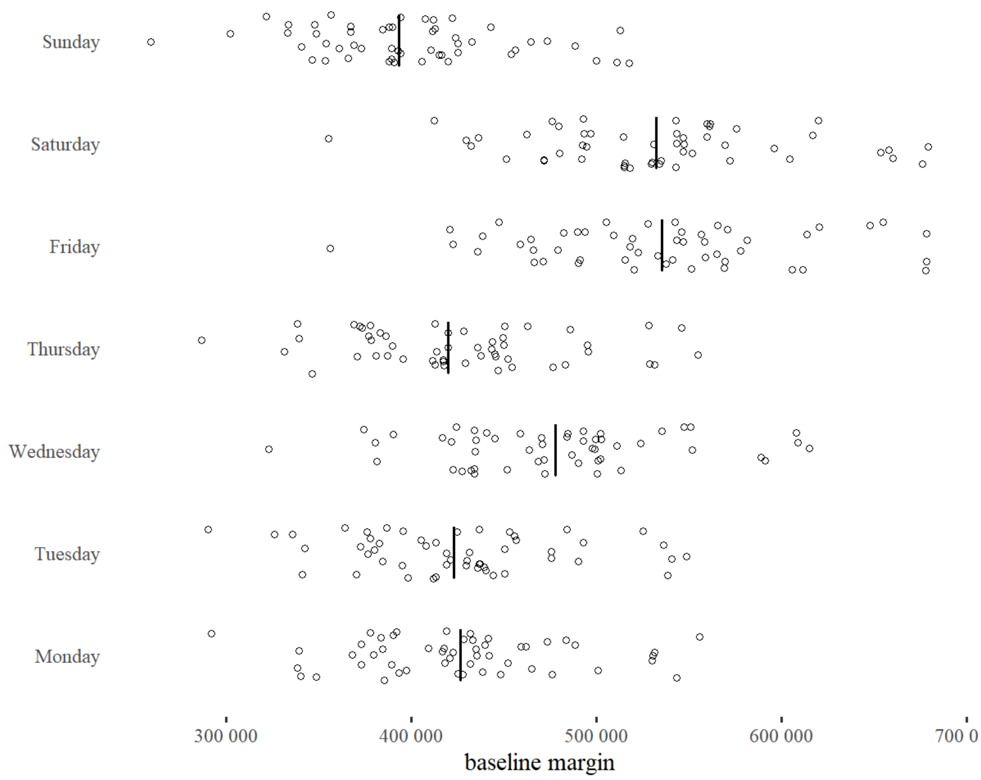


Figure 5.3. Baseline margins of each weekday. Each point represent one observation and each line the median of the points. On average Wednesday, Friday and Saturday appear to performing better than the other weekdays.

some seasonality patterns in categories. Most of the categories (e.g. 1, 2 and 3) seem to peak during Christmas although some exceptions exist (e.g. 6 and 10). These patterns help us to time promotions correctly.

We can go even deeper to the product level. In Figure 5.5, we can see the sparklines of the products in Category 1. We can again see that promotion and media activity cause sudden rises in margins, and we can see when each product has been promoted. We can also notice that performance is not uniform across the products within the category. For example, Product 1 performs much better than other products in the same category. We are also able to make similar findings within other categories with the help of their sparklines (see Appendix A.2).

In summary, the detailed information about performance and seasonalities helps us find the most effective products and to time their promotions correctly, something which we could not have done with a higher-level analysis. However, we have to spend more time analysing the data and making decisions. We could find even more improvement opportunities by analysing, for example, the optimal promotion intervals and discount levels but our approach is not able to produce information about decay rates, and it does not have enough different price levels for each product to produce a reliable estimate. Overall, many fine-tuning opportunities exist on the product level.

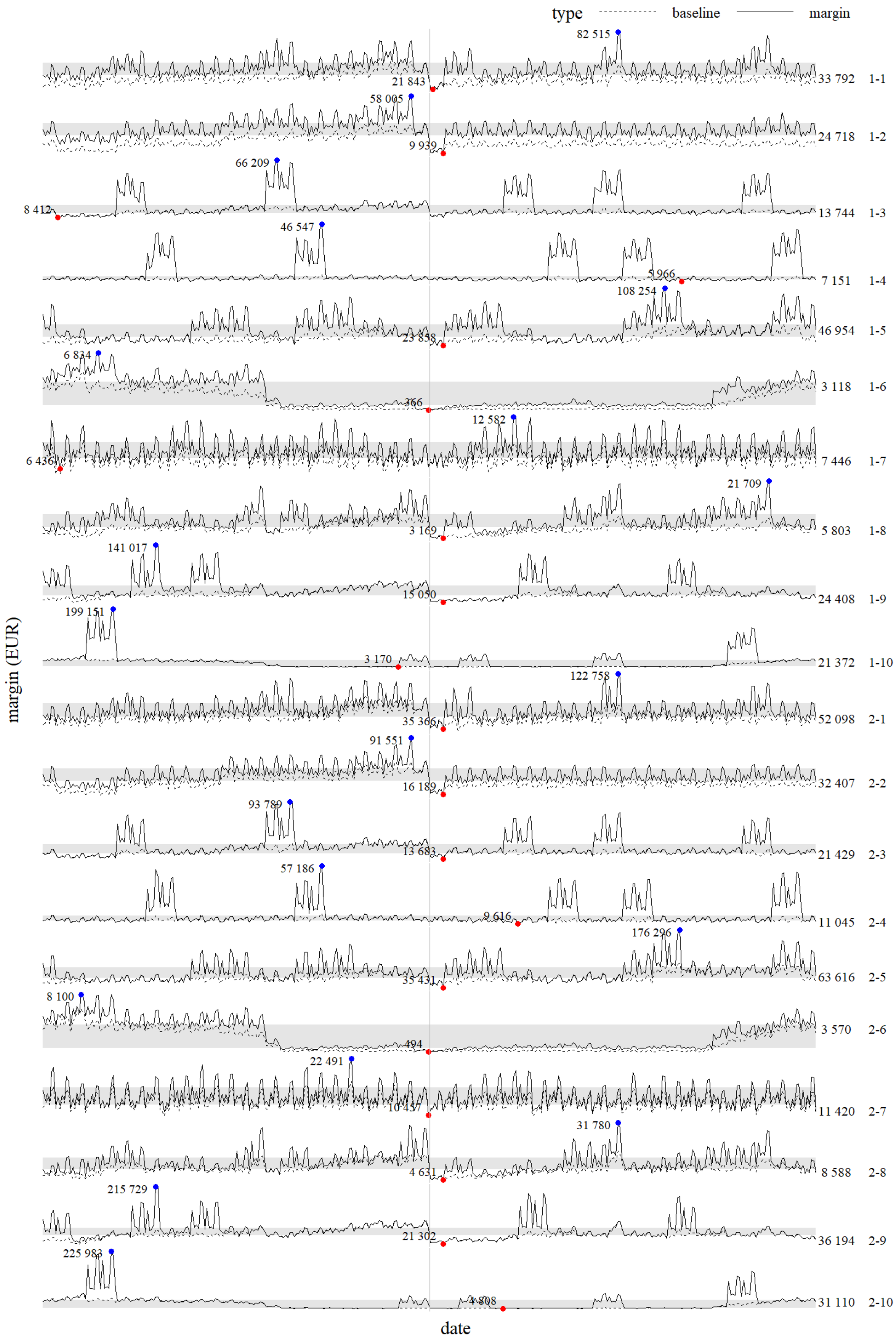


Figure 5.4. Category sparklines over one year. The blue point represents the maximum, the red point the minimum, the grey area the 25%–75% value range and the number right of the line the last value of the series.

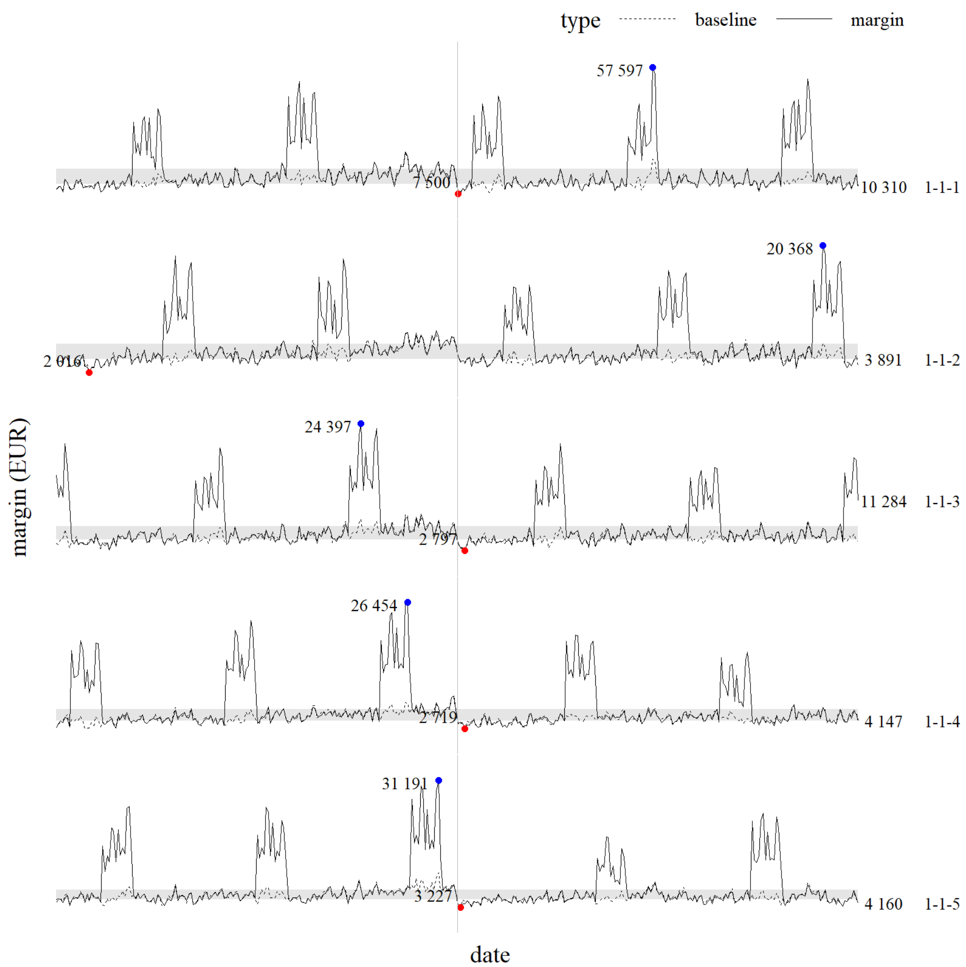


Figure 5.5. Sparklines for the products in Category 1 and Region 1.

5.2 Profit, media spend and average ROAS

We can use the baseline decomposition to calculate the incremental profits from advertising and thus, also the ROAS numbers. Table 5.1 shows the total margin alongside incremental margins, media spending, total profit and ROAS numbers. We can see that on the national level paid search was more effective than TV (ROAS 2.2 vs 1.3). The total margin was EUR 220 million, the TV spending EUR 15 million and the search spending EUR 14.3 million, amounting to EUR 190 million in profit (EUR 12.26 per capita). Total media spending was EUR 29.3 million, which is about 13.4% of the total margin. On the regional level, we can again see that TV is more effective in Region 2 (1.7 vs 1.0) and paid search in Region 1 (2.4 vs 1.9), even though spending is quite even across the regions. Region 2 also brings in more profit (€116 mill. vs €74 mill.) probably because of the higher population.

Figure 5.6 shows a quick overview of the ROAS figures. We can see that Categories 1, 2, 3, 5, 9 and 10 have a high marketing efficiency. In turn, Categories 4, 6, 7 and 8 are disappointing. These categories mainly have low margin products, whereas the best-performing ones mostly have high margin products. Performance is also not uniform within categories. Some products have much higher performance than others. For example, Product 1 in Category 9 performs much better than other products in the same category. This detailed information enables us to focus the advertising budgets on the most effective products. However, we have to do increasingly more analysis and fine-tuning.

Table 5.1. Overview of the total performance. All figures except ROAS in thousands ('000).

	Comb.	Tot. margin	Incr. margin TV	Incr. margin Search	TV spend	Search spend	Tot. profit	TV ROAS	Search ROAS
Nation	1	219 316	20 072	31 001	15 000	14 333	189 983	1.3	2.2
Region	1	88 355	7 590	17 159	7 500	7 016	73 839	1.0	2.4
	2	130 961	12 482	13 842	7 500	7 317	116 144	1.7	1.9
Region Category	1 1	16 096	1 197	2 690	750	735	14 611	1.6	3.7
	1 2	11 001	933	2 926	750	735	9 516	1.2	4.0
	1 3	7 285	529	1 265	750	735	5 800	0.7	1.7
	1 4	4 430	359	1 070	750	716	2 964	0.5	1.5
	1 5	17 707	1 492	2 802	750	735	16 222	2.0	3.8
	1 6	793	43	249	750	619	-576	0.1	0.4
	1 7	3 163	142	229	750	735	1 678	0.2	0.3
	1 8	3 372	422	638	750	718	1 904	0.6	0.9
	1 9	16 204	1 458	2 553	750	735	14 719	1.9	3.5
	1 10	8 303	1 014	2 736	750	552	7 001	1.4	5.0
	2 1	24 355	2 141	2 280	750	735	22 870	2.9	3.1
	2 2	15 676	1 392	2 409	750	735	14 191	1.9	3.3
	2 3	11 345	877	1 035	750	735	9 860	1.2	1.4
	2 4	6 141	600	761	750	735	4 656	0.8	1.0
	2 5	27 622	2 554	2 376	750	735	26 137	3.4	3.2
	2 6	947	79	193	750	734	-537	0.1	0.3
	2 7	5 328	256	190	750	735	3 843	0.3	0.3
	2 8	4 759	653	487	750	735	3 274	0.9	0.7
	2 9	24 712	2 517	2 018	750	735	23 227	3.4	2.7
	2 10	10 076	1 414	2 094	750	703	8 623	1.9	3.0
Region Category Product	1 1 1	6 097	480	823	150	147	5 800	3.2	5.6
	1 1 2	2 163	163	365	150	147	1 866	1.1	2.5
	1 1 3	2 530	160	461	150	147	2 233	1.1	3.1
	1 1 4	2 453	184	504	150	147	2 156	1.2	3.4
	1 1 5	2 853	210	536	150	147	2 556	1.4	3.6
	1 2 6	2 304	205	600	150	147	2 007	1.4	4.1
	1 2 7	2 150	185	608	150	147	1 853	1.2	4.1
	1 2 8	2 207	171	602	150	147	1 910	1.1	4.1
	1 2 9	1 970	158	548	150	147	1 673	1.1	3.7

Number	Tot. margin	Incr. margin TV	Incr. margin Search	TV spend	Search spend	Tot. profit	TV ROAS	Search ROAS
1 2 10	2 370	213	569	150	147	2 073	1.4	3.9
1 3 11	2 289	150	278	150	147	1 992	1.0	1.9
1 3 12	1 877	126	321	150	147	1 580	0.8	2.2
1 3 13	1 620	121	285	150	147	1 323	0.8	1.9
1 3 14	819	81	202	150	147	522	0.5	1.4
1 3 15	680	51	179	150	147	383	0.3	1.2
1 4 16	1 107	74	230	150	147	810	0.5	1.6
1 4 17	315	31	119	150	128	37	0.2	0.9
1 4 18	799	138	177	150	147	502	0.9	1.2
1 4 19	747	24	213	150	147	450	0.2	1.4
1 4 20	1 463	92	331	150	147	1 166	0.6	2.3
1 5 21	3 980	317	722	150	147	3 683	2.1	4.9
1 5 22	3 546	300	447	150	147	3 249	2.0	3.0
1 5 23	3 943	407	630	150	147	3 646	2.7	4.3
1 5 24	3 486	374	546	150	147	3 189	2.5	3.7
1 5 25	2 752	93	457	150	147	2 455	0.6	3.1
1 6 26	259	14	88	150	131	-22	0.1	0.7
1 6 27	161	9	50	150	119	-108	0.1	0.4
1 6 28	144	8	45	150	113	-119	0.1	0.4
1 6 29	121	7	35	150	126	-155	-	0.3
1 6 30	109	5	30	150	131	-172	-	0.2
1 7 31	346	22	30	150	147	49	0.1	0.2
1 7 32	688	25	31	150	147	391	0.2	0.2
1 7 33	661	48	66	150	147	364	0.3	0.4
1 7 34	931	32	64	150	147	634	0.2	0.4
1 7 35	538	15	39	150	147	241	0.1	0.3
1 8 36	1 217	139	209	150	147	920	0.9	1.4
1 8 37	356	59	83	150	139	67	0.4	0.6
1 8 38	387	64	54	150	147	90	0.4	0.4
1 8 39	187	19	68	150	138	-101	0.1	0.5
1 8 40	1 224	142	225	150	147	927	0.9	1.5
1 9 41	8 570	888	1 305	150	147	8 273	5.9	8.9
1 9 42	2 143	97	470	150	147	1 846	0.6	3.2
1 9 43	3 344	333	461	150	147	3 047	2.2	3.1
1 9 44	605	33	119	150	147	308	0.2	0.8
1 9 45	1 541	107	198	150	147	1 244	0.7	1.3
1 10 46	1 445	256	384	150	129	1 166	1.7	3.0
1 10 47	1 944	246	754	150	90	1 704	1.6	8.4
1 10 48	1 294	163	471	150	93	1 051	1.1	5.1
1 10 49	1 548	194	598	150	94	1 304	1.3	6.4
1 10 50	2 071	156	529	150	145	1 776	1.0	3.7
2 1 1	9 762	809	732	150	147	9 465	5.4	5.0
2 1 2	3 205	287	307	150	147	2 908	1.9	2.1
2 1 3	3 732	294	378	150	147	3 435	2.0	2.6
2 1 4	3 487	368	403	150	147	3 190	2.5	2.7
2 1 5	4 168	382	461	150	147	3 871	2.5	3.1
2 2 6	3 341	300	494	150	147	3 044	2.0	3.4
2 2 7	3 021	292	507	150	147	2 724	1.9	3.5
2 2 8	3 089	242	504	150	147	2 792	1.6	3.4

Number	Tot. margin	Incr. margin TV	Incr. margin Search	TV spend	Search spend	Tot. profit	TV ROAS	Search ROAS
2 2 9	2 780	243	444	150	147	2 483	1.6	3.0
2 2 10	3 445	316	460	150	147	3 148	2.1	3.1
2 3 11	3 733	253	233	150	147	3 436	1.7	1.6
2 3 12	2 954	220	263	150	147	2 657	1.5	1.8
2 3 13	2 532	178	237	150	147	2 235	1.2	1.6
2 3 14	1 186	124	164	150	147	889	0.8	1.1
2 3 15	940	101	139	150	147	643	0.7	0.9
2 4 16	1 610	134	187	150	147	1 313	0.9	1.3
2 4 17	408	45	106	150	147	111	0.3	0.7
2 4 18	1 102	242	125	150	147	805	1.6	0.8
2 4 19	937	47	138	150	147	640	0.3	0.9
2 4 20	2 084	132	206	150	147	1 787	0.9	1.4
2 5 21	6 053	539	593	150	147	5 756	3.6	4.0
2 5 22	5 601	545	405	150	147	5 304	3.6	2.8
2 5 23	6 213	663	561	150	147	5 916	4.4	3.8
2 5 24	5 443	625	454	150	147	5 146	4.2	3.1
2 5 25	4 312	181	363	150	147	4 015	1.2	2.5
2 6 26	304	27	65	150	147	7	0.2	0.4
2 6 27	192	16	39	150	147	-105	0.1	0.3
2 6 28	173	13	36	150	147	-124	0.1	0.2
2 6 29	147	12	28	150	147	-150	0.1	0.2
2 6 30	131	11	24	150	147	-166	0.1	0.2
2 7 31	566	39	23	150	147	269	0.3	0.2
2 7 32	1 206	46	28	150	147	909	0.3	0.2
2 7 33	1 093	83	55	150	147	796	0.6	0.4
2 7 34	1 564	61	55	150	147	1 267	0.4	0.4
2 7 35	899	26	30	150	147	602	0.2	0.2
2 8 36	1 887	237	167	150	147	1 590	1.6	1.1
2 8 37	457	85	63	150	147	160	0.6	0.4
2 8 38	529	95	43	150	147	232	0.6	0.3
2 8 39	237	28	58	150	147	-60	0.2	0.4
2 8 40	1 649	208	156	150	147	1 352	1.4	1.1
2 9 41	13 450	1 546	1 061	150	147	13 153	10.3	7.2
2 9 42	2 828	155	359	150	147	2 531	1.0	2.4
2 9 43	5 181	558	355	150	147	4 884	3.7	2.4
2 9 44	765	68	82	150	147	468	0.5	0.6
2 9 45	2 487	191	161	150	147	2 190	1.3	1.1
2 10 46	1 697	346	264	150	147	1 400	2.3	1.8
2 10 47	2 234	325	558	150	134	1 950	2.2	4.2
2 10 48	1 502	246	359	150	137	1 215	1.6	2.6
2 10 49	1 768	271	432	150	138	1 480	1.8	3.1
2 10 50	2 876	226	479	150	147	2 579	1.5	3.3

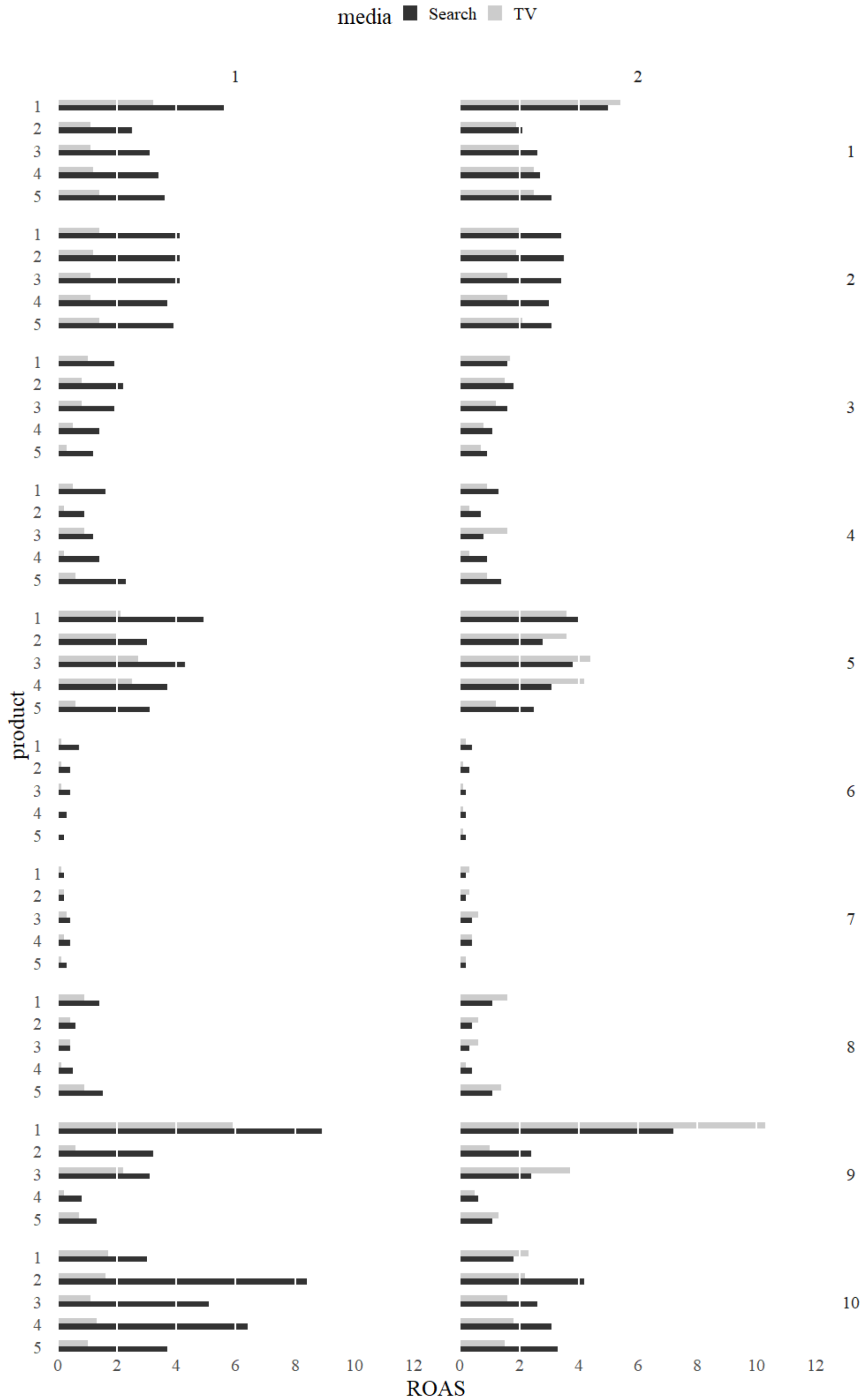


Figure 5.6. Product-level marketing performance.

5.3 Recommendations and new results

Now that we have analysed the marketing system and performance, we can make improvement suggestions at various levels.

Basic weekly aggregated MMM Based on the ROAS numbers, it is better to focus on paid search rather than TV. Because we do not have an estimate of the S-curve, we decide to modestly decrease the TV budget by around 20% and move the freed funds to the paid search budget. This move gives us a TV budget of EUR 12 million and a paid search budget of EUR 17.7 million. The total budget remains the same. We focus budgets on seasons where media has been the most efficient: May and the high peaks seen in the spring and autumn. In turn, we reduce budgets from seasons when media has been ineffective: the Christmas period and July. To ensure that this new allocation increases media effectiveness, we keep the original promotion and media tables the same. If the underlying marketing activity changed, the new allocation might not produce good results.

Geo-category model With the geo-category model, we similarly adjust the media budgets by moving 20% of the TV budget to paid search and keep the total budget the same. However, we also make regional and category adjustments. With TV, we allocate around 60% of the budget to Region 2 and the rest to Region 1. With paid search, we do the opposite. We also focus the budget on the most efficient categories. The best ones are 1, 5 and 9, and they will receive around 55% of the budget. The second best are 2, 3 and 10, and they will receive around 40% of the budget. Categories 4 and 8 receive the rest of the budget and the remaining Categories 6 and 7 receive nothing. Categories 6 and 7 are low margin categories and trying to fix them would be difficult. Discounts are halted in the cut categories.

We also adjust media and promo timings of categories. We ensure that categories are in promotion during their high season. The categories are also promoted more often thanks to increased budget. The promotional pattern between products inside a category remains the same. We also ensure that discounts always support media activity. Finally, we adjust media timings and run TV ads on Wednesday, Friday and Saturday instead of the typical Monday, Friday and Saturday.

We also put 70% of the weekly budget on Friday and Saturday and the rest on Wednesday.

Product level model With the promotion model, we make the same media and regional adjustments, adjust TV timings and keep the total budget the same. However, instead of categories, we focus the budget on the best products. The star products are Products 1-1, 5-1, 5-3, 5-4, 9-1 and 10-2 and the focus in budgeting is on them. The budgets of products in Categories 6 and 7 and of Products 4-2, 8-2, 8-3, 8-4 and 9-4 are cut. We increase the advertising frequency of the remaining products again. We also focus media budgets on the top seasons of individual products and ensure that discounts always support media activity. We do not analyse or change discount rates because of the limited pricing data.

After simulating the additional year, we calculated the change in total profit and marketing efficiency. In Table 5.2, we can see the overall profits in the original and each improvement scenario. The table also lists the TV and paid search ROAS figures. As we can see, the basic marketing mix model was only able to provide a slight improvement. The improvement opportunities available at its granularity level raised the overall profit only by 1.7% or EUR 1.2 million. Also, the paid search ROAS remained unchanged, and the TV ROAS only increased by 7.7%. It seems that allocating media budgets to times where media had been the most effective was not a successful strategy, most likely because it also increased the budgets of poor performing categories during those periods.

Table 5.2. Results from the last year of the simulation. Total media spending is equal in all scenarios (approx. €29 mill.).

Scenario	**Profit ’000 000**	**+%**	**TV ROAS**	**+%**	**Search ROAS**	**+%**
Original	188.7	-	1.3	-	2.1	-
Basic MMM	191.9	1.7%	1.4	7.7%	2.1	0%
Geo-category model	229.5	21.6%	2.8	115.4%	3.3	57.1%
Product model	251.1	33.1%	4.0	207.7%	3.8	81.0%

The geo-category model with daily data, on the other hand, was much more successful, producing a 21.6% (€40.8 mill.) increase in profit, a 115.4% increase in the TV ROAS and a 57.1% increase in the paid search ROAS. Thanks to the granularity, we were able to make categorical and regional adjustments and focus the budgets on the best performing categories and to use media where it was most effective. We were also able to put promotions where they were the most effective. Moreover, we were able to optimise the TV timings.

The best results were, however, reached with the product-level model. The improvements provided a 33.1% increase (€62.3 mill.) in profits, a 207.7% improvement in the TV ROAS and an 81.0% increase in the paid search ROAS. With product level granularity, we were able to reveal performance differences within categories and to nail budget proportions between products. We were also able to make some product-level timing optimisations. Overall, the product-level data gave us the best understanding and control over the marketing environment. However, at the same time, the decision-making complexity increased, and we had to make much more decisions than on higher levels, even though we did not consider all types of optimisations, such as product pricing, advertising frequency and interproduct effects such as cannibalisation and halo. With a broader product range, even our simple optimisation approach would have quickly become too complicated to do manually, and we would have had to implement some improvement finding system. For large retailers, building such a system is consequently likely to become a necessity in order to efficiently find improvement opportunities and to integrate them into the processes of the company.

6. Limitations

We can analyse the limitations in the two areas of our study: the literature review and the experiment.

The literature review discussed and presented a wide variety of literature related to marketing effectiveness, marketing mix modelling and the retail industry. We were able to identify the various impacts of marketing, especially in the retail industry. We also analysed the key marketing measure, the ROMI, and described how the typical approach to measuring it, the marketing mix modelling, was limited in accuracy and largely unadjusted to meet the needs of the retail marketing environment. Based on these findings, we were able to suggest several improvements that would make the marketing mix modelling more suitable for the retail environment. Although we did not much consider the practical implementation of these improvements, e.g. by writing mathematical equations, we were able to provide a useful summary of the directions toward which companies and researchers could develop marketing mix modelling. We were also able to identify various impacts of the improvements.

The literature review also led to our experiment, where we tested the impact of the main improvement: improved model granularity. We used simulation to isolate the impact of the granularity by eliminating the impact of model selection and accuracy. Simulation also allowed us to accurately calculate the true underlying incremental effects in the data, something which we could not have ensured in a real-life experiment. In the simulation, we saw the best performance increases when making improvements at the highest granularity, which supported our hypothesis that improved modelling granularity leads to more detailed and effective improvement opportunities. We were also able to demonstrate some of the various analysis opportunities that are available at the different model granularity levels. We did not, however, take into account some detail adjustments, including advertising intervals, pricing and interproduct effects such as cannibalisation and halo. These would have likely increased the importance of modelling granularity as they can be analysed most accurately with product-level data.

The main question related to our experiment is whether the results from our simulation are generalisable for the retail industry. That generalisability depends on whether the simulation reflects a retail environment well. Our simulation was a simplified marketing environment with two media, two regions, ten categories and 50 products. The simulation attempted to imitate realistic marketing and promotion activity and produced a reasonable sales time series with reasonable marketing performance. Our region–category–product hierarchy reflected the increasing variance in performance similarly to Figure 3.1. Moreover, we took advantage of the Google AMSS by Zhang and Vaver (2017) that is likely the most advanced publicly available marketing system simulator. It takes into account various effects of the retail marketing environment, such as the S-curve, adstock and competitor activity, and enabled us to adjust the marketing system to match the retail environment as well as we could.

The main difference between our simulation and a real marketing environment, however, comes from the small product, media and region count. Although this simplified marketing environment helped us keep the simulation simple and finding improvement opportunities manageable, it does not reflect the environment of the biggest of retailers but rather a portion of it or the environment of a small retailer. The hypothesis is, thus, likely to hold for most retailers. Moreover, the improvement opportunities are likely to be higher for large retailers because of the higher variance and complexity in the marketing environment, although finding the opportunities will require more work and automation.

Despite finding support for the hypothesis, the exact numbers produced by our experiment are unlikely to hold for all retailers; the nature of the marketing environment and the level of initial optimisation are unlikely to be same for all retailers. The numbers are also derived in a simulation where we assume that the marketer is fully accurately able to measure the impact of marketing. However, in practice, accurate measurement is difficult, even when applying the most advanced models. For example, on the product level, it is likely impossible to register all the marketing efforts and market trends that affect the sales. Consequently, retailers may not be able to fully discover and capitalise on the available improvement opportunities and the impact of different modelling granularities may vary. Despite these limitations, the figures act as solid support for the hypothesis and consequently, for the importance of modelling granularity in the retail marketing environment. A problem for retailers is whether they can take advantage of that granularity.

7. Conclusions

Having now done a comprehensive literature review and studied how improved modelling granularity can support retail marketing, we can now answer the research questions.

1. What limitations do typical marketing mix models have, especially in the retail marketing environment?

Typical marketing mix modelling is limited in reliability and is largely unadjusted for the retail marketing environment. The main limitation comes from the lack of granular data. A typical MMM data set is monthly or weekly aggregated on a national level. This aggregation limits the scope of the model to the same level, creating a tyranny of average where information about the underlying performance structure is lost. The lost information, in turn, reduces opportunities in improving marketing performance, for example, through promotion management and geographical strategies. Despite limited information, a high-level model is not entirely useless. It can help in adjusting marketing budgets and act as a sanity check for spending levels, although, as we saw in the experiment, the results are unlikely to be as effective as with more granular models. Moreover, a high-level analysis can lead to wrong conclusions because of the hidden internal performance structure. For example, in our experiment, Christmas appeared weak, mostly because of the unattractive products promoted during it, which led us to shift budgets away from it, although we could have fixed the situation with a better promotion mix. To avoid similar errors, managers should approach such high-level analyses carefully and attempt to understand what is behind the performance of each marketing activity.

The lack of data also limits the reliability of marketing mix models. In a highly aggregated data set, the number of data points per parameter becomes low and the variability in data is reduced, making fitting more unreliable and causing extrapolation uncertainty. The lack of reliable data, for example, about competitor activity, can also introduce inaccuracies or force modellers to omit variables.

Model forms are also limited in several ways. Firstly, marketing mix models often suffer from biases and unverified assumptions about the nature of the marketing environment. The lack of proper model validation exacerbates the situation. Secondly, the side effects of promotion activity (cannibalisation, stock-up, halo), which are especially relevant for the retail industry, are not often integrated together into promotion models in the literature. These effects can have a significant impact on promotion performance and should be measured. Moreover, models often do not attempt to split incremental sales between the promotion and media. Thirdly, MMM often only analyses short-term sales and disregards the long-term effects. Marketing can have brand-building and brand-eroding effects that are reflected in the baseline, and which marketers should take into account in marketing performance measurement. Finally, MMM produces results of past performance but which are used to make predictions. The lack of forecasting can make these results misleading.

2. How can marketing mix modelling be improved for the retail marketing environment through:

1. *better data,*
2. *better models, and*
3. *model validation through simulation?*

Based on the identified limitations, we can formulate several improvements in the three areas:

Data In terms of data, the best way for retailers to improve marketing mix modelling is to increase data granularity. Retailers can increase granularity in at least three dimensions: frequency, geography and product hierarchy. The improvement would not only increase data quantity and variability and also enable identifying pockets of growth and better performing areas in, for example, regions and products. Data granularity must, however, be improved uniformly because a low granularity in one data type can limit the granularity of the whole model. Besides increasing granularity, retailers could improve data quality in areas where it has been unreliable: for example, in competitor activity and marketing spending. Higher quality, in turn, could increase model accuracy and ensure that managers have the correct information about the state of marketing. Table 3.1 gives a summary of the improvements in data.

Models The best way to adjust marketing mix models for the retail industry is to build models that support both high-level and bottom-level decision making. One approach is to start with a bottom-level product model and then aggregate the results upwards to yield high-level media, area and category results. This approach provides consistent results across the organisation while still preserving the information about the underlying performance. The complete information about the performance structure, in turn, supports managers in making more efficient decisions about marketing on all levels. However, as we saw in the experiment, increased granularity also increases the amount of information rapidly and consequently, the complexity of decision making. Thus, automated decision making systems and dashboards may be needed to support managers.

Besides granularity, the model form should be improved to counter biases and to account for more effects. For example, model developers should find ways to counter various biases, such as the selection bias. Similarly, model developers should adjust models to account for the side effects of promotion (cannibalisation, stock-up, halo), which influence heavily in the retail marketing environment and be able to differentiate between the effect of the promotion and the media. Furthermore, models should draw a baseline that separates the impact of marketing on the consumer mindset in the long-term, thus enabling retailers to analyse the long-term impact of marketing. Similarly, loyalty program data can help retailers to study the effect on marketing on individual customers and customer segments and to develop more targeted marketing. Finally, retailers could develop forecasting methods to predict future performance, e.g., for promotions to optimise inventory levels and vendor compensation or for marketing strategies to better understand the impact of changes.

Creating a granular model that integrates all these effects is a complex task. Bayesian modelling, however, could support the development of complex models through its flexibility, ability to use priors for parameters and the ability to report on parameter and model uncertainty. Some of the needed improvement areas are also quite unstudied in the literature. Thus, companies must make investments in model development. Table 3.3 highlights the improvement opportunities in model development.

Model validation Without proper testing of model performance, it is difficult to rely on models in decision making. Models can have various unproven assumptions about the marketing environment and suffer from various biases. Just getting a proper fit is not enough to ensure correct results. Model developers can use simulations to test models flexibly and cost-effectively under different marketing environments, for example, under different levels of seasonality or funnel effects, to discover in what areas they perform well or poorly. This testing supports the iterative development of models and increases managers trust in them. Simulators can also be developed further to reflect new changes or findings of the marketing environment and combined with other validation methods to guarantee the result quality. Although developers may never be able to develop a wholly accurate simulation, simulation offers a promising path for model development. Furthermore, companies planning to purchase MMM services can take advantage of simulations to compare suppliers' models.

3. How can such improvements support retail marketing management?

We can categorise the improvement opportunities as follows: improving model granularity, comprehensive effect measurement, improving data quality and model validation through simulation. In our experiment, we focused on the impact of the first improvement area. The others, we can analyse with the support of our literature review.

Improving model granularity As we saw in the experiment, improving data and model granularity allowed us to investigate the whole performance structure of retail marketing. We were able to analyse the performance of each product over time, measure uplifts from promotions and split the uplifts between the different media and the promotion itself. These numbers could be aggregated upwards to calculate product, category, regional and total performance, also as ROAS figures. The results allowed us to make changes at different levels of granularity. On the product level, we made changes to the promotion mix, allocating more funds to products, regions and media with higher performance. We also timed promotions better and focused TV on the most productive days. On the category-region level, we made changes to category, region and media allocations, category promotion timings and TV airing dates. Finally, on the top-level, we only made changes between advertising channels and moved budgets towards seasons when media had been effective.

As a result, we saw the best improvements with the most granular model, the product model with daily data. The total profit increased by 33.1%, and we were able to triple the TV ROAS and almost double the paid search ROAS. The basic marketing mix model with national and weekly granularity was only able to increase total profit by 1.7% and the TV ROAS by 7.7% while the paid search efficiency remained unchanged. Although the numbers are likely to differ between retailers, they act as a support for the hypothesis of the experiment: increased modelling granularity provides more detailed and effective improvement opportunities in the retail marketing environment. The improvement opportunities are also likely to be higher with large retailers that have more complexity and variance in their performance structure. Similarly, taking advantage of all the possible optimisations that become available on the lower levels, such as discount or advertising interval optimisations, would provide even bigger improvements. Overall, improved granularity gives retailers a better understanding of their marketing performance structure and enables making more accurate and effective decisions.

Comprehensive effect modelling Promotions can have various side effects, such as cannibalisation and halo. Without modelling such impacts, a retailer could unknowingly promote products that attract bargain-hunters and stockpilers and miss the star products that bring in the largest purchase baskets. Marketing can also have significant long-term effects on consumer behaviour. Through comprehensive modelling, managers can understand the true impact of promotions and make better decisions. Furthermore, if they can integrate forecasting into their models, they can make better predictions, for example, about needed stock levels and required vendor compensation. Finally, model developers need to correctly model various other effects in the marketing environment. These can be, for example, funnel effects from other media to paid search, which could otherwise distort ROAS figures. Modelling such effects accurately will help managers get more reliable results.

Improving data quality High data quality is necessary for accurate results. If the modeller feeds inaccurate data into the model, it will only spit out inaccurate data. The impact depends on the quantity of inaccurate data, the scale of data errors and the model form. Gathering quality data is also essential for forming control over marketing. As noted in the example of ICA AB, large amounts of marketing spending were hidden and thus, out of control of the top marketing managers. Forming a complete picture helps, in turn, making better decisions.

Model validation through simulation Simulation supports testing and improving the reliability of models. The marketing mix models are used to make rather significant decisions, and if a model is unvalidated, it is hard to rely on in the decision making. Simulation enables testing of model performance under different market conditions and supports their iterative development. Retailers can also test the quality of external MMM services with the simulations. The ultimate goal is to produce reliable figures for the retailer.

Overall, the improvements would bring significant benefits for retailers through more granular, complete and trustworthy analyses. These analyses would give retailers a better understanding of marketing and enable them to make more effective decisions. However, to take advantage of these opportunities, retailers need to invest in the development of marketing mix modelling. The problem areas of MMM, their improvement opportunities and the ultimate impacts are highlighted in Table 7.1. Overall, our study was successful in answering all the research questions through the literature review and the simulation. The results enable us to formulate a road map for retail marketing mix modelling in the next chapter, thus meeting our research objective.

Table 7.1. Problems and improvement opportunities of marketing mix modelling.

Area	Problem	Improvement	Impact
Data	Low data granularity limits scope and data quantity.	Improve data granularity (frequency, geography, product hierarchy).	Better model accuracy and understanding of the underlying performance structure
	Lack of accurate data limits model accuracy.	Invest in gathering quality data, especially in key areas (sales and marketing activity).	Better model accuracy
Models	Limited in scope	Deepen modelling granularity as data gathering improves.	Better model accuracy and understanding of the underlying performance structure
	Most models do not model promotion side effects, long-term effects and future performance	Attempt to capture effects with new model forms.	Better understanding of promotion, long-term and future performance
	Many sources of biases	Tackle biases with new models forms and better data.	Better model accuracy
Model valida-tion	Model accuracy is not often tested properly.	Use realistic simulations for iterative testing.	Effective model development, better understanding of model performance
	Comparing the quality of different MMM services is difficult.	Use simulators to compare supplier performance.	Easier to pick the best MMM service provider

8. Recommendations

As called for by managers and academics, marketers should invest in the measurement of marketing effectiveness. It gives them credibility by making their actions measurable and enables them to gain back trust and influence from the top management. Moreover, measuring marketing effectiveness supports marketers in improving their performance. The more granular and complete level of analysis they can reach in their measurement, the better the results they can achieve. Many retailers, however, might not have the resources to jump into measuring marketing effectiveness on a full scale. They can lack, for example, data collection processes, modelling know-how and funds for investment. Therefore, companies may want to develop their marketing effectiveness measurement capabilities gradually over time.

Figure 8.1 represents a road map for developing marketing mix modelling in the retail environment. It shows the three different models, which we analysed in the experiment, as checkpoints on the way.

Basic MMM is a simple starting point for development. It offers a simple way to measure media performance and saturation. Moreover, a proper uplift–baseline decomposition can support in the quantification of long-term effects. Data collection and decision making at this level are also more straightforward than on lower levels. However, the benefits are likely to be smaller as well. Managers must also be aware of possible errors from the low amount of data. Furthermore, if the underlying distribution of promotions and regions or other activities is not uniform, it can make a channel look superficially more effective than the other. For example, if the retailer promotes more lucrative items in one media than in another one.

Geo-category models offer the logical next step. These allow making much more granular decisions than the basic MMM by guiding re-

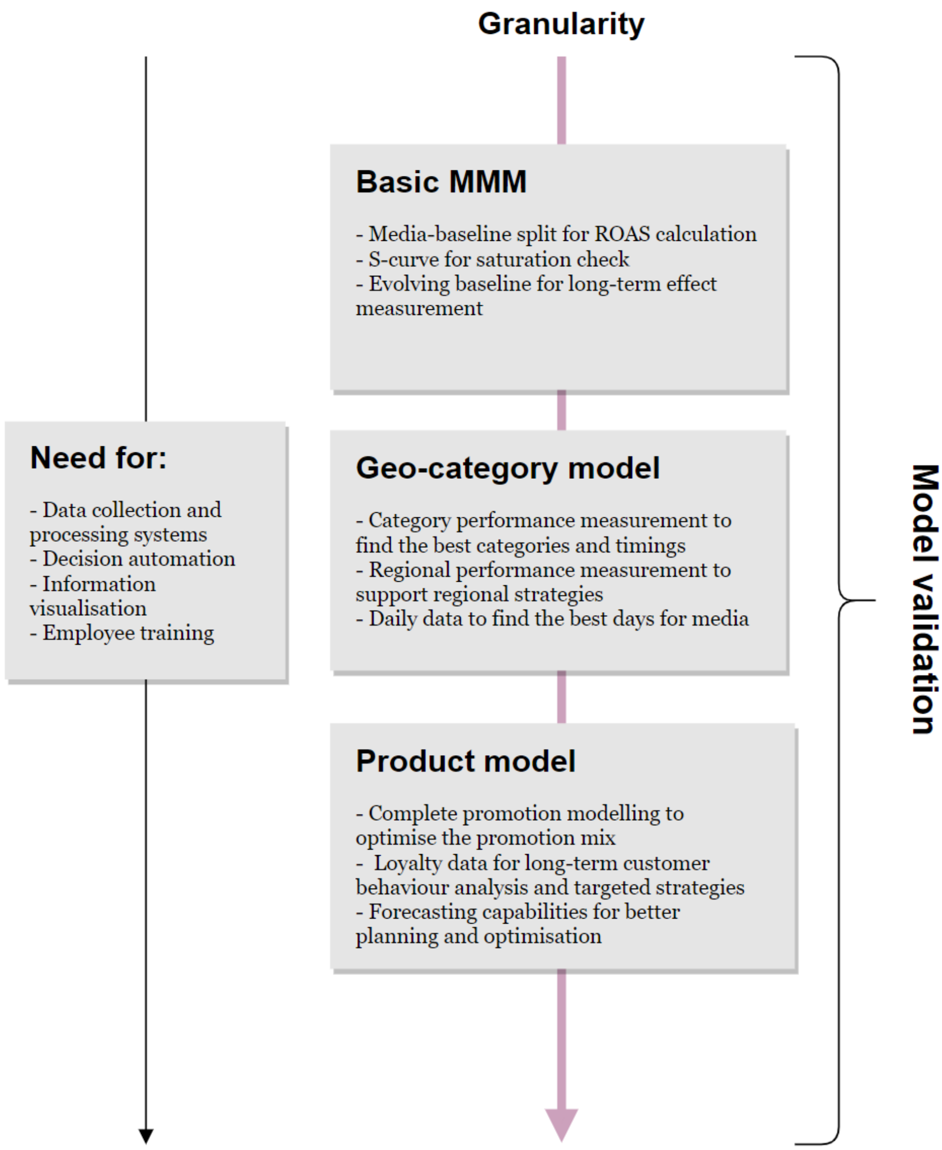


Figure 8.1. A road map for developing marketing mix modelling for the retail marketing environment.

tailers to focus budgets on the most productive regions and product categories. With daily data collection, retailers can also analyse intraweek patterns and focus media on the productive days. Moving on to the level will require investments into data collection, but the benefits are likely to be more significant. Moreover, the model accuracy is likely to be higher thanks to the higher data quantity and variability.

Product models are the ultimate goal. On this level, retailers should attempt to measure promotions and their side effects to form a complete understanding of promotion performance. Also, by providing split between the effects of the media and promotion, retailers can accurately calculate the effect of the media. The results can also be aggregated higher to form a uniform understanding of the performance structure across the organisation. At the product level, retailers can improve long-term effect measurement with loyalty program data and create targeted strategies for customer segments. Forecasting capabilities may also turn out to be useful, for example, estimating correct stock-levels and optimising promotion mixes. Reaching this level will require even more investments in data collection and processing than at higher levels. However, the benefits are likely to be the most significant.

Model development from the basic MMM to the product level model must be supported by model validation. If model accuracy is unverified, it difficult for the retailer to rely on it. Various biases and false assumptions may affect model accuracy. With simulation, retailers can iteratively and flexibly test and develop their models. Simulators, however, need to be developed to together with the models and supported with traditional model testing methods.

Retailers also need to adjust their processes to take full advantage of marketing effectiveness measurement. A granular model can produce so much information that it overwhelms managers. If managers and other employees are not effectively able to incorporate the results into the workflow of the company, nothing will change. Interactive dashboards and automated decision-making systems can support this integration. Actively updated dashboards are always available and support decision making just when it is needed. Automated decision-making systems, in turn, can ease the workload of managers by automating marketing decisions. Besides systems, retailers should invest in educating their employees. As noted earlier, the ROMI and other marketing measures often misunderstood, and thus, the systems may not automatically lead to better results. Therefore, retailers need to ensure that their managers and employees understand the essential metrics and know how to use them. Overall, retailers may need to make large changes in their organisations.

To avoid developing measurement capabilities by themselves, retailers can purchase ready services offered by consultancies and software service providers. These services can be very advanced and produce quality results. Yet, companies must approach such companies carefully. If the company does not have granular and high-quality data, the results from such services are likely to be limited and possibly unreliable. Thus, companies may need to work on improving their data quantity and quality. They also may need to work with the supplier to ensure an automated data exchange. These changes can require additional investments. Besides ensuring the data are available, companies should preferably test or demand proof of the quality of the offered service. The companies service may be untested with simulations and produce unreliable results. Thus, the purchasing company should test the service beforehand with realistic simulations or require assurance from the provider in other ways.

Overall, marketing mix modelling offers significant opportunities for retailers, which they should not ignore.

Bibliography

Magid M. Abraham and Leonard M. Lodish. Promoter: An Automated Promotion Evaluation System. *Marketing Science*, 6(2):101–123, 5 1987. ISSN 0732-2399. doi: 10.1287/mksc.6.2.101. URL http://pubsonline.informs.org/doi/abs/10.1287/mksc.6.2.101.

Arthur W. Allaway, Richard M. Gooner, David Berkowitz, and Lenita Davis. Deriving and exploring behavior segments within a retail loyalty card program. *European Journal of Marketing*, 40(11/12):1317–1339, 11 2006. ISSN 0309-0566. doi: 10.1108/03090560610702830. URL https://www.emeraldinsight.com/doi/10.1108/03090560610702830.

Tim Ambler and John Roberts. Beware the silver metric: marketing performance measurement has to be multidimensional. Technical report, London Business School, Centre for Marketing, London, United Kingdom, 2005. URL http://facultyresearch.london.edu/docs/Beware_the_Silver_Metric_05-709.pdf.

Thomas Bauer and Sascha Lehmann. Budget Prioritization. In Jesko Perrey and Dennis Spillecke, editors, *Retail Marketing and Branding*, pages 129–146. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition, 10 2015. doi: 10.1002/9781119207900.ch9. URL http://doi.wiley.com/10.1002/9781119207900.ch9.

Gabriele Bavagnoli, Clarisse Magnin-Mallez, Tarek Mansour, and Stefano Zerbi. Fact-based Promotion Management. In Jesko Perrey and Dennis Spillecke, editors, *Retail Marketing and Branding*, chapter 15, pages 231–245. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition, 10 2015. doi: 10.1002/9781119207900.ch15. URL http://doi.wiley.com/10.1002/9781119207900.ch15.

Keith Becker, Jim Sprigg, and Alex Cosmas. Estimating individual promotional campaign impacts through Bayesian inference. *Journal of Consumer Marketing*, 31(6/7):541–552, 11 2014. ISSN 0736-3761. doi: 10.1108/JCM-06-2014-1006. URL http://www.emeraldinsight.com/doi/10.1108/JCM-06-2014-1006.

Neil T. Bendle and Charan K. Bagga. The Metrics That Marketers Muddle. *MIT Sloan Management Review*, 57(3):73–82, 2016. URL http://mitsmr.com/1LojUzZ.

Les Binet and Peter Field. *The Long and the Short of it: Balancing Short and Long-Term Marketing Strategies*. Institute of Practitioners in Advertising, 2013. ISBN 085294134X. URL https://barbanouille.files.wordpress.com/2016/07/the_long_and_short_of_it_pdf_doc.pdf.

Cadent Consulting Group. 2017 Marketing Spending Industry Study. Technical report, Cadent Consulting Group, 2017. URL http://cadentcg.com/wp-content/uploads/2017-Marketing-Spending-Study.pdf.

Peter M. Cain. Limitations of conventional marketing mix modelling. *Admap Magazine*, (493):48–51, 4 2008. URL https://www.warc.com/content/paywall/article/limitations_of_conventional_marketmix_modelling/88108.

Peter M. Cain. Marketing Mix Modelling and Return on Investment. In Philip J Kitchen, editor, *Integrated Brand Marketing and Measuring Returns*, chapter 4, pages 94–130. Palgrave Macmillan UK, London, 1st edition, 2010. ISBN 9781403998613. URL http://link.springer.com/10.1057/9780230297340_4.

David Chan and Michael Perry. Challenges And Opportunities In Media Mix Modeling. Technical report, Google Inc, 2017. URL https://ai.google/research/pubs/pub45998.

Aiyou Chen, David Chan, Mike Perry, Yuxue Jin, Yunting Sun, Yueqing Wang, and Jim Koehler. Bias Correction For Paid Search In Media Mix Modeling. Technical report, Google Inc, 2018. URL https://ai.google/research/pubs/pub46861.

Lee G. Cooper, Penny Baron, Wayne Levy, Michael Swisher, and Paris Gogos. PromoCast™: A New Forecasting Method for Promotion Planning. *Marketing Science*, 18(3):301–316, 8 1999. ISSN 0732-2399. doi: 10.1287/mksc.18.3.301. URL http://pubsonline.informs.org/doi/abs/10.1287/mksc.18.3.301.

Dentsu Aegis Network Ltd. Global Ad Spend Forecasts. Technical Report February, 2019. URL https://www.dentsuaegisnetwork.com/reports/ad_spend_february_2019_asset.

Tim Eales. Price and Promotion in Western Economies. Technical report, Information Resources, Inc. (IRI)., 2016. URL https://www.iriworldwide.com/IRI/media/IRI-Clients/International/Price-Promotion-in-Western-Economies-a-Pause-in-Promotion-Escalation.pdf.

Facebook. Conversion Lift: Helping Marketers Better Understand the Impact of Facebook Ads, 2015. URL https://www.facebook.com/business/news/conversion-lift-measurement.

Paul Farris, Dominique M. Hanssens, James Lenskold, and David Reibstein. Marketing return on investment: Seeking clarity for concept and measurement. *Applied Marketing Analytics*, 1(3):267–282, 2015. URL https://www.ingentaconnect.com/content/hsp/ama/2015/00000001/00000003/art00010.

Fournaise Group. 80% of CEOs Do Not Really Trust Marketers (Except If They Are "ROI Marketers®"), 2012. URL https://www.fournaisegroup.com/ceos-do-not-trust-marketers/.

Google. Brand Lift, 2018. URL https://www.thinkwithgoogle.com/products/brand-lift/.

Dominique M. Hanssens and Koen H. Pauwels. Demonstrating the Value of Marketing. *Journal of Marketing*, 80(6):173–190, 11 2016. ISSN 0022-2429. doi: 10.1509/jm.15.0417. URL http://journals.sagepub.com/doi/10.1509/jm.15.0417.

Olli Herrala. Evaluating cannibalization between items in retail promotions. Bachelor's thesis, Aalto University, 2018. URL http://urn.fi/URN:NBN:fi:aalto-201810235501.

Christian Homburg, Arnd Vomberg, Margit Enke, and Philipp H. Grimm. The loss of the marketing department's influence: is it really happening? And why worry? *Journal of the Academy of Marketing Science*, 43(1):1–13, 1 2015. ISSN 0092-0703. doi: 10.1007/s11747-014-0416-3. URL http://link.springer.com/10.1007/s11747-014-0416-3.

Harald Hruschka, Martin Lukanowicz, and Christian Buchta. Cross-category sales promotion effects. *Journal of Retailing and Consumer Services*, 6(2): 99–105, 4 1999. ISSN 09696989. doi: 10.1016/S0969-6989(98)00026-5. URL https://linkinghub.elsevier.com/retrieve/pii/S0969698998000265.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge, 1st edition, 4 2015. ISBN 9780521885881. doi: 10.1017/CBO9781139025751. URL https://www.cambridge.org/core/product/identifier/9781139025751/type/book.

Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, and Jim Koehler. Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. Technical report, Google Inc, 2017. URL https://ai.google.com/research/pubs/pub46001.

Tobias Karmann, Lars Köster, and Thomas Meyer. Budget Sizing: The Million-dollar Question. In Jesko Perrey and Dennis Spillecke, editors, *Retail Marketing and Branding*, chapter 8, pages 117–128. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition, 10 2015. doi: 10.1002/9781119207900.ch8. URL http://doi.wiley.com/10.1002/9781119207900.ch8.

Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1):140–181, 2 2009. ISSN 1384-5810. doi: 10.1007/s10618-008-0114-1. URL http://link.springer.com/10.1007/s10618-008-0114-1.

Peter S.H. Leeflang, Jaap E Wieringa, Tammo H.A. Bijmolt, and Koen H Pauwels. *Modeling Markets*, volume 1 of *International Series in Quantitative Marketing*. Springer New York, New York, NY, 2015. ISBN 978-1-4939-2085-3. doi: 10.1007/978-1-4939-2086-0. URL http://link.springer.com/10.1007/978-1-4939-2086-0.

Randall A. Lewis and Justin M. Rao. The Unfavorable Economics of Measuring the Returns to Advertising. *The Quarterly Journal of Economics*, 130(4):1941–1973, 11 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv023. URL httsp://doi.org/10.1093/qje/qjv023.

Y. Jackie Luan and K. Sudhir. Forecasting Marketing-Mix Responsiveness for New Products. *Journal of Marketing Research*, 47(3):444–457, 6 2010. ISSN 0022-2437. doi: 10.1509/jmkr.47.3.444. URL http://journals.sagepub.com/doi/10.1509/jmkr.47.3.444.

Max Magni, Anne Martinez, Rukhshana Motiwala, and Alex Rodriguez. Are US consumers feeling greater financial security?, 2017. URL https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/are-us-consumers-feeling-greater-financial-security.

Puneet Manchanda, Asim Ansari, and Sunil Gupta. The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions. *Marketing Science*, 18(2):95–114, 5 1999. ISSN 0732-2399. doi: 10.1287/mksc.18.2.95. URL http://pubsonline.informs.org/doi/abs/10.1287/mksc.18.2.95.

Alberto La Mantia. Optimization of Econometric Models using Evolutionary Algorithms. Master's thesis, University of Copenhagen, 2015.

Chiara Mauri. Card loyalty. A new emerging issue in grocery retailing. *Journal of Retailing and Consumer Services*, 10(1):13–25, 1 2003. ISSN 09696989. doi: 10.1016/S0969-6989(02)00036-X. URL `https://linkinghub.elsevier.com/retrieve/pii/S096969890200036X`.

Leigh McAlister, Edward I. George, and Yung-Hsin Chien. A Basket-mix Model to Identify Cherry-picked Brands. *Journal of Retailing*, 85(4):425–436, 12 2009. ISSN 00224359. doi: 10.1016/j.jretai.2009.09.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S002243590900061X`.

Ecaterina Mhitarean-Cuvsinov. Marketing Mix Modelling from multiple regression perspective. Master's thesis, KTH Royal Institute of Technology, 2017. URL `http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-208474`.

Christine Moorman. The CMO Survey: Highlights & Insights Report. Technical Report February, World Market Watch LLC, Durham, North Carolina, 2019. URL `https://cmosurvey.org/results/february-2019/`.

Martin Natter, Thomas Reutterer, Andreas Mild, and Alfred Taudes. Practice Prize Report —An Assortmentwide Decision-Support System for Dynamic Pricing and Promotion Planning in DIY Retailing. *Marketing Science*, 26(4):576–583, 7 2007. ISSN 0732-2399. doi: 10.1287/mksc.1060.0217. URL `http://pubsonline.informs.org/doi/10.1287/mksc.1060.0217`.

Koen Pauwels. Truly Accountable Marketing: The Right Metrics for the Right Results. *GfK Marketing Intelligence Review*, 7(1):8–15, 5 2015. ISSN 1865-5866. doi: 10.1515/gfkmir-2015-0001. URL `http://content.sciendo.com/view/journals/gfkmir/7/1/article-p8.xml`.

Koen Pauwels and Dave Reibstein. Challenges in Measuring Return on Marketing Investment. In *Review of Marketing Research*, pages 107–124. Routledge, 7 2017. ISBN 0309056091094. doi: 10.4324/9781315088754-5. URL `https://www.taylorfrancis.com/books/9781351551052/chapters/10.4324/9781315088754-5`.

Hean Lee Poh and Teo Jašić. Forecasting and analysis of marketing data using neural networks: a case of advertising and promotion impact. In *Proceedings the 11th Conference on Artificial Intelligence for Applications*, volume 7, pages 224–230. IEEE Comput. Soc. Press, 12 1998. ISBN 0-8186-7070-3. doi: 10.1109/CAIA.1995.378819. URL `http://ieeexplore.ieee.org/document/378819/`.

Minna Rantanen and Päivi Kauppinen. Mediamainonnan määrä kääntyi nousuun vuonna 2018. Technical report, Kantar TNS Oy, Espoo, 2019. URL `https://www.kantar.fi/sites/default/files/lehdistotiedote_mediamainonnan_maara_2018.pdf`.

Roland T. Rust, Tim Ambler, Gregory S. Carpenter, V. Kumar, and Rajendra K. Srivastava. Measuring Marketing Productivity: Current Knowledge and Future Directions. *Journal of Marketing*, 68(4):76–89, 10 2004. ISSN 0022-2429. doi: 10.1509/jmkg.68.4.76.42721. URL `http://journals.sagepub.com/doi/10.1509/jmkg.68.4.76.42721`.

David A. Schwartz. Measuring the Effectiveness of Your Company's Advertising. *Journal of Marketing*, 33(2):20, 4 1969. ISSN 00222429. doi: 10.2307/1249397. URL `https://www.jstor.org/stable/1249397?origin=crossref`.

Jorge M. Silva-Risso, Randolph E. Bucklin, and Donald G. Morrison. A Decision Support System for Planning Manufacturers' Sales Promotion Calendars. *Marketing Science*, 18(3):274–300, 8 1999. ISSN 0732-2399. doi: 10.1287/mksc.18.3.274. URL `http://pubsonline.informs.org/doi/abs/10.1287/mksc.18.3.274`.

David W. Stewart. Marketing accountability: Linking marketing actions to financial results. *Journal of Business Research*, 62(6):636–643, 6 2009. ISSN 01482963. doi: 10.1016/j.jbusres.2008.02.005. URL `https://linkinghub.elsevier.com/retrieve/pii/S014829630800074X`.

Yunting Sun, Yueqing Wang, Yuxue Jin, David Chan, and Jim Koehler. Geo-level Bayesian Hierarchical Media Mix Modeling. Technical report, Google Inc, 2017. URL `https://ai.google/research/pubs/pub46000`.

Tuomas Viitanen. Considering cannibalization and halo effects to improve demand forecasts, 2018. URL `https://www.relexsolutions.com/cannibalization-halo-effects-in-demand-forecasts/`.

Yueqing Wang, Yuxue Jin, Yunting Sun, David Chan, and Jim Koehler. A Hierarchical Bayesian Approach to Improve Media Mix Models Using Category Data. Technical report, Google Inc, 2017. URL `https://ai.google/research/pubs/pub45999`.

Kimberly A. Whitler and Neil Morgan. Why CMOs never last. *Harvard Business Review*, 94(4):46–54, 2017. URL `https://hbr.org/2017/07/the-trouble-with-cmos`.

Stephanie Zhang and Jon Vaver. Introduction to the Aggregate Marketing System Simulator. Technical report, Google Inc, 2017. URL `https://ai.google/research/pubs/pub45996`.

A. Appendices

A.1 Aggregate Marketing System Simulator

Consumer states and migration

The Aggregate Marketing System Simulator by Zhang and Vaver (2017) is designed to simulate natural customer behaviour and the influence of marketing interventions on them. Simply put, it does this by dividing the population into various state combinations and then simulating movements of users between them. Overall, the simulator uses six different dimensions for states. The three first ones relate to customers' relationship with the product category:

- **Market state** specifies whether the consumer has an interest in the product. A customer can be either 'in-market' or 'out-of-market'.

- **Satiation state** defines whether a recent purchase has satiated the consumer. A consumer can either be 'satiated' or 'unsatiated'.

- **Activity state** describes where the consumer is along the path to making a purchase. A consumer can either be in 'inactive', 'exploratory' or 'purchase' state. In the two latter states, the consumer must also be 'in-market' and 'unsatiated'.

The three last ones, in turn, relate to the consumers' brand relationships:

- **Brand favourability state** describes the consumers' awareness of the advertiser's brand. A consumer can either be 'unaware', 'negative', 'neutral', 'somewhat favourable', or 'favourable'.

- **Brand loyalty state** represents consumers' loyalty to the brand. A consumer can either be 'switcher', 'loyal', or 'competitor-loyal'.

- **Brand availability state** describes the physical availability of the product, e.g., does the consumer shop in stores selling the product? A consumer can either be in 'low', 'average' or 'high' availability state.

Table 1.1. Category and brand states of the AMSS.

State type	Potential values
Product	
Market	in-market, out-of-market
Satiation	satiated, unsatiated
Activity	inactive, exploratory, purchase
Brand	
Favourability	unaware, unfavourable, neutral
	somewhat favourable, favourable
Loyalty	switcher, loyal, competitor-loyal
Availability	low, average, high

Table 1.1 summarises the all the dimensions and their states. The natural behaviour of the market and marketing interventions move customers between these states based on stochastic matrices set by the user. These movements are calculated during each interval (e.g. each week or day) by a sequence of events. Each event (e.g. TV advertising) takes the current population segmentation as input and returns an updated population segmentation to the next event in the sequence, along with related output variables (e.g. media spend and volume). The final event of each interval is the sales event, where consumers in the purchase state make their purchases either from the advertiser or its competitors. Figure 1.1 portrays this whole flow of events over time. The events in the sequence are standardised and for media, two different modules exist:

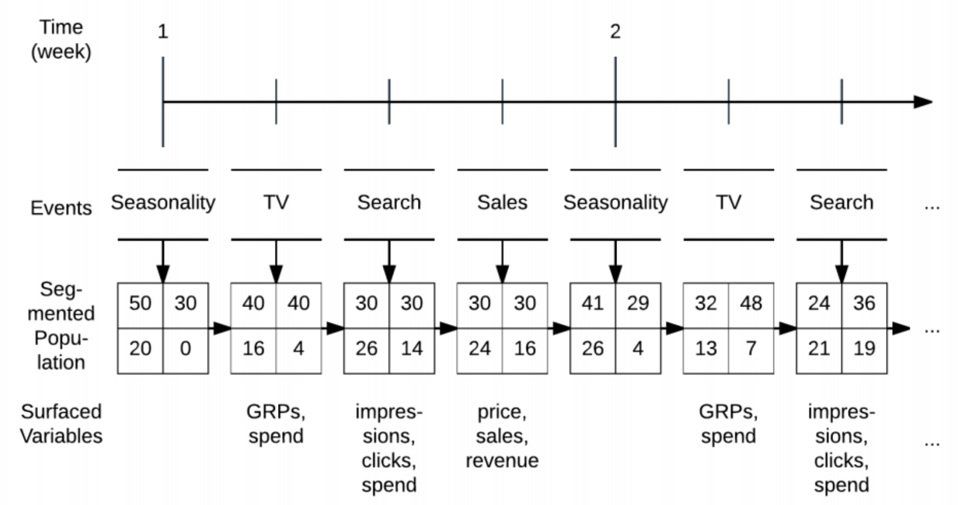


Figure 1.1. Overview of the simulator flow (Zhang and Vaver, 2017).

Traditional media module

Traditional media module is a module for simulating simple media channels such as the TV. The modeller can change various parameters such as audience size and composition, media volume and spend and media effectiveness. In short, the flow of the module is the following:

1. **Calculate the media audience**. The size of the audience for the media channel is determined by the probabilities set by the user for each state type. E.g. 'a person in the exploratory state has a 30% chance of being in the audience'.

2. **Calculate the weekly spend**. The budget of each period is distributed between the media flightings based on a pattern specified by the modeller. E.g. 'run TV commercials only every other week on Wednesdays'.

3. **Calculate media volume**. The media volume, e.g. the number of exposures, is determined by the media budget and unit cost. The modeller can alter the cost and its variance.

4. **Reach and frequency**. The average frequency, i.e. the number of ad exposures among consumers with at least one exposure, is calculated. It will be used in the next phase to determine the impact of media activity.

5. **Update the population segmentation**. The population states are transformed using the transition matrix set by the modeller. The proba-

bilities are scaled using the average frequencies in a Hill transformation, resulting in an S-curve in media effectiveness. The transformation enables modelling initially accelerating and eventually diminishing returns for media channels.

The overall flow of the module with its inputs and variables is displayed in Figure 1.2.

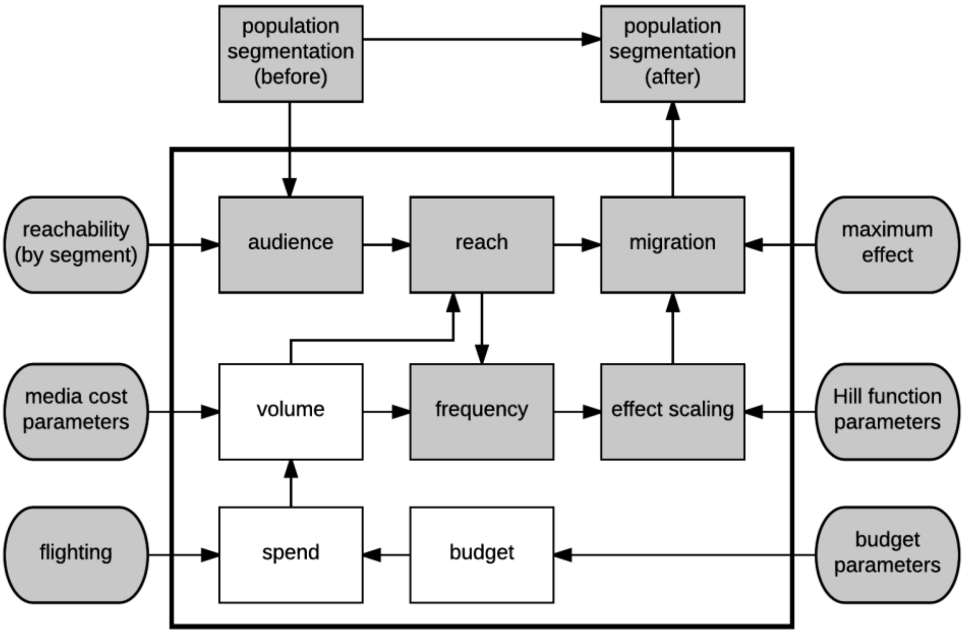


Figure 1.2. The flow of the traditional media module. The ovals represent the input parameters of the module, whereas the rectangles represent either hidden (shaded) or observed (unshaded) variables (Zhang and Vaver, 2017).

Paid search media module

The paid search media module simulates Google-like paid search, which is auction-based and thus limited in inventory. The module simulates a wide range of variables of the system, including bidding, keyword lists, paid clicks and impressions. The overall flow of the module is the following:

1. **Determine the campaign settings from the budget**. The modeller can impact the campaign spending by altering either the spending cap, size of the bid or/and keyword list length.

2. **Calculate query volume**. The modeller can define the probabilities and rates at which different segments search for the products.

3. **Calculate impressions, clicks and spend**. The system calculates the number of impressions (ads shown the user) and then uses the click-through rate (CTR) to calculate the number of clicks generated by those impressions. Customers from different segments have different CTRs. Finally, the spend is calculated by multiplying the number of clicks with the cost-per-click (CPC).

4. **Update the population segmentation**. A consumer may have different types of exposures to search: a consumer may either (a) see no paid ads from the advertiser, resulting in an organic experience, (b) see the advertiser's paid ad but not click on it or (c) see the paid ad an click on it. These three exposure types have different levels of effect, which are used to scale the typical transition matrices. The modeller can change the relative effectiveness of these exposure types to affect the impact of each search type.

The overall flow of the module with its inputs and variables is displayed in Figure 1.3.

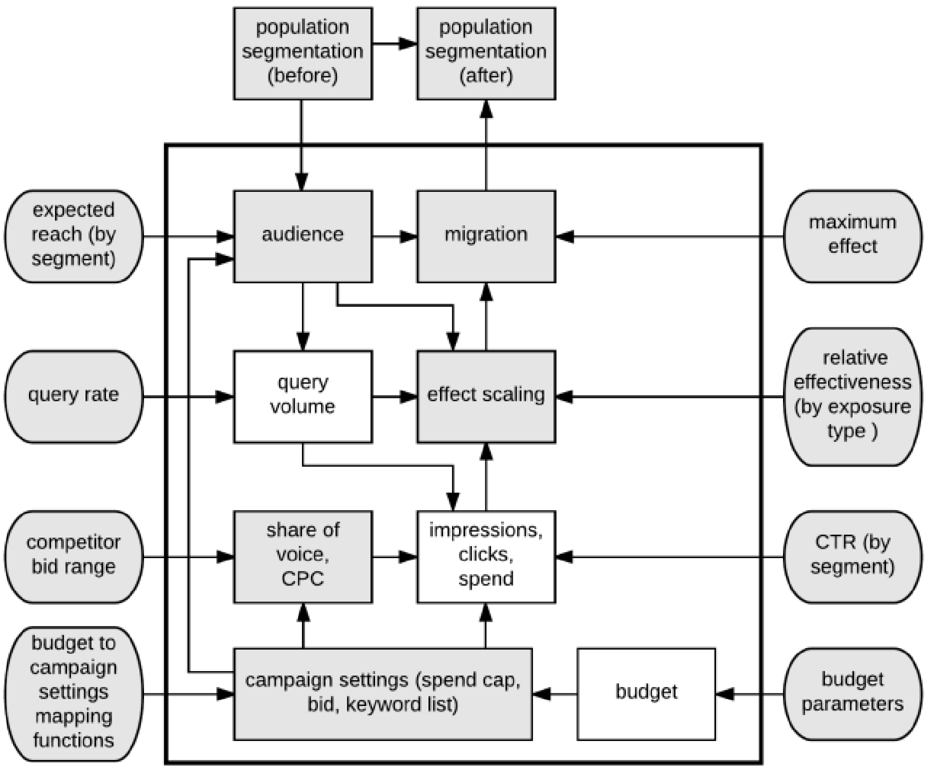


Figure 1.3. The flow of the paid search media module. The ovals represent the input parameters of the module, whereas the rectangles represent either hidden (shaded) or observed (unshaded) variables (Zhang and Vaver, 2017).

Sales event

The final event in the sequence is the sales event. During the event, the advertiser's and the competitor's sales per consumer segment are calculated for that time interval. In short, a linear demand curve of price determines the probability of purchase in each of the advertiser's customer segment. The y-intercept α of the curve determines the probability of purchase at the price point of 0. The slope β determines the decrease in probability with the increase in price. The demand curve in Figure 1.4 represents these curves in various segments in a competitor free environment.

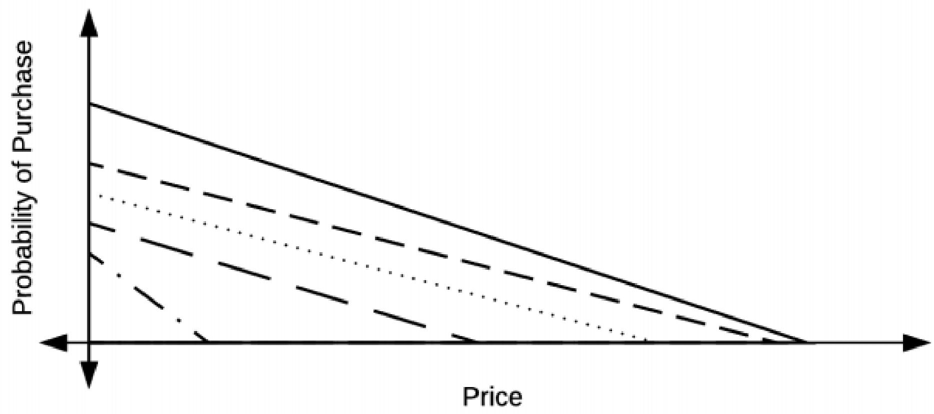


Figure 1.4. Demand curves of various customer segments in a competitor free environment (Zhang and Vaver, 2017).

The advertiser's demand curves are affected by competitor behaviour, which is simplified by another linear demand curve. The curve's y-intercept γ determines the probability that a consumer will purchase a competitor's product when the advertiser's products price is too high to make any sales. The slope ω, in turn, defines at what rate the advertiser's and competitor's sales replace each other. When the slope is 1, the competitor's sales are unaffected by the competitor. Typically, competitor-loyal consumers can behave in this way. Similarly, if the slope is 0, the consumers are advertiser-loyal and the advertiser's sales are unaffected by the competitor. People who have no brand preference, i.e. 'switchers', can be set to use the slope of 0.5. These slopes and their interaction with the advertiser's demand curves are illustrated in more detail in Figure 1.5.

Finally, we need to define the average number of products purchased by the consumer and its slope with the price. After that, we can calculate the sales. First, the simulator uses the demand curves to divide customers into those who will not buy, those who will buy from the competitor and

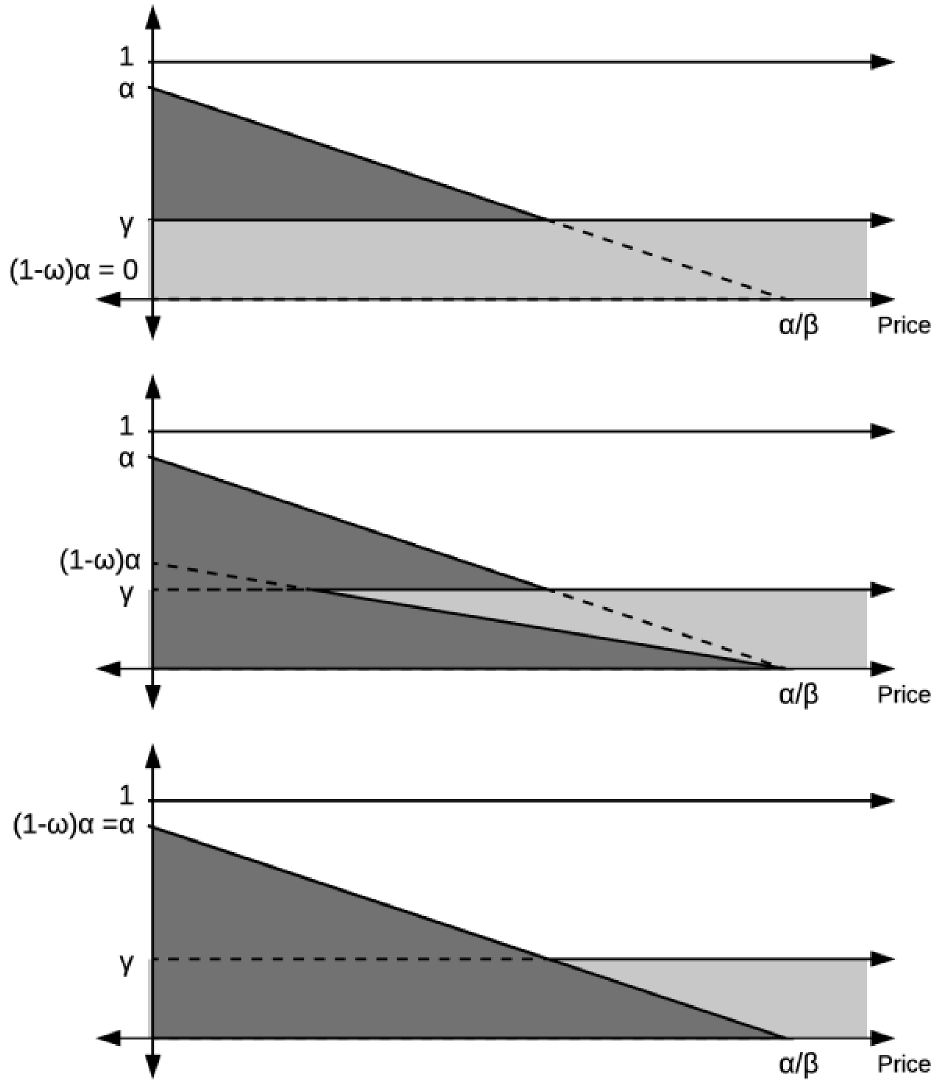


Figure 1.5. Demand curves in a competitive situation. At any price p, the height of the lightly shared region is the probability for a purchase of the competitor's brand. The combined height of the darker shade in turn is the probability for a purchase of the advertiser's brand. The three figures represent the effect of different replacement slopes ω: $\omega = 1$ (top), $\omega = 1/2$ (middle) and $\omega = 0$ (bottom). (Zhang and Vaver, 2017).

those who will buy from the advertiser. Then, using the average number of purchases and the price of the product, the simulator generates sales. After purchase, the users become satiated and will make no purchases directly afterwards. The user can adjust the time needed for the consumers to become unsatiated again. The purchase experience can also be set to affect the consumers' relationship with the brand.

Ground truth

To calculate the ROAS figures needed for MMM testing, the difference between the current media strategy and its counterfactual has to be estimated. This can be done by generating data sets with and without spending in a certain media channel and comparing them. More formally, if we have a marketing strategy $b = (b_m)_{1:M}$ where b_m is the budget for the m-th media channel. Let $m = 1$ represent television. First, we generate N_1 data sets $D_{n_1}(b)$, where $n_1 = 1, \ldots, N_1$ and the media budget stays the same. Then, we generate the counterfactual strategy b', where no budget is assigned to TV, i.e., $b'_1 = 0$. Again, we have N_2 data sets $D_{n_2}(b')$, where $n_2 = 1, \ldots, N_2$. Finally, we compare the differences in TV spend x_{n_i} against the differences in revenue y_{n_i} in the data sets.

$$\hat{\theta} = \frac{1}{N_1 N_2} \sum \frac{y_{n_1}(b) - y_{n_2}(b)}{x_{n_1}(b) - x_{n_2}(b)} \tag{1.1}$$

By the law of large numbers, the $\hat{\theta}$ approaches the true ROAS, θ. The accuracy of the estimate can be calculated as a margin of error from the variability in the sample. Both the estimation of ROAS and its error are incorporated in a single function in the AMSS.

More information and details on the simulator can be found on the AMSS GitHub page at `https://github.com/google/amss` and in the introductory article by Zhang and Vaver (2017).

A.2 Product-level sparklines

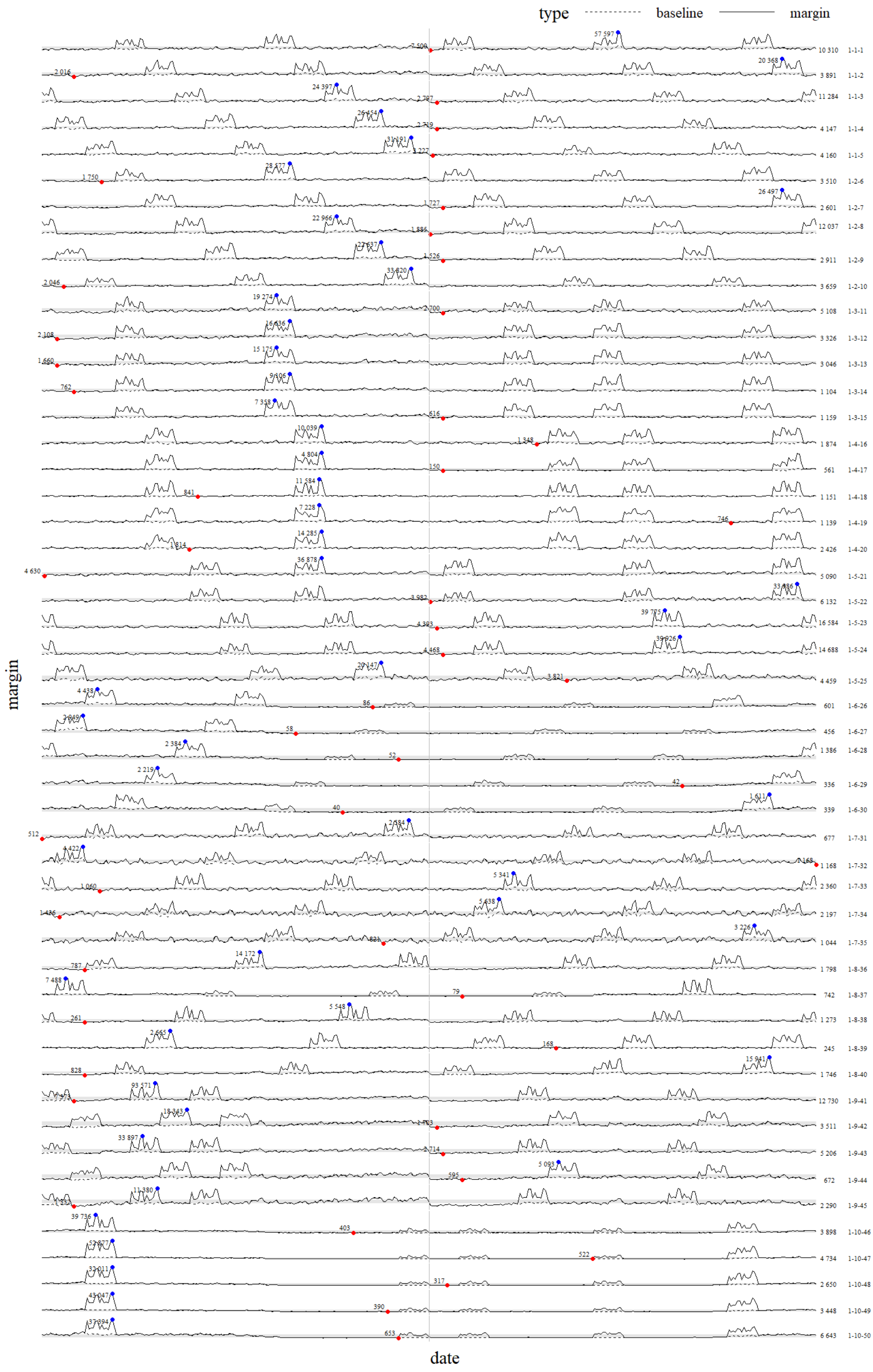


Figure 1.6. Region 1 – Daily product sparklines of the base data set over one year.

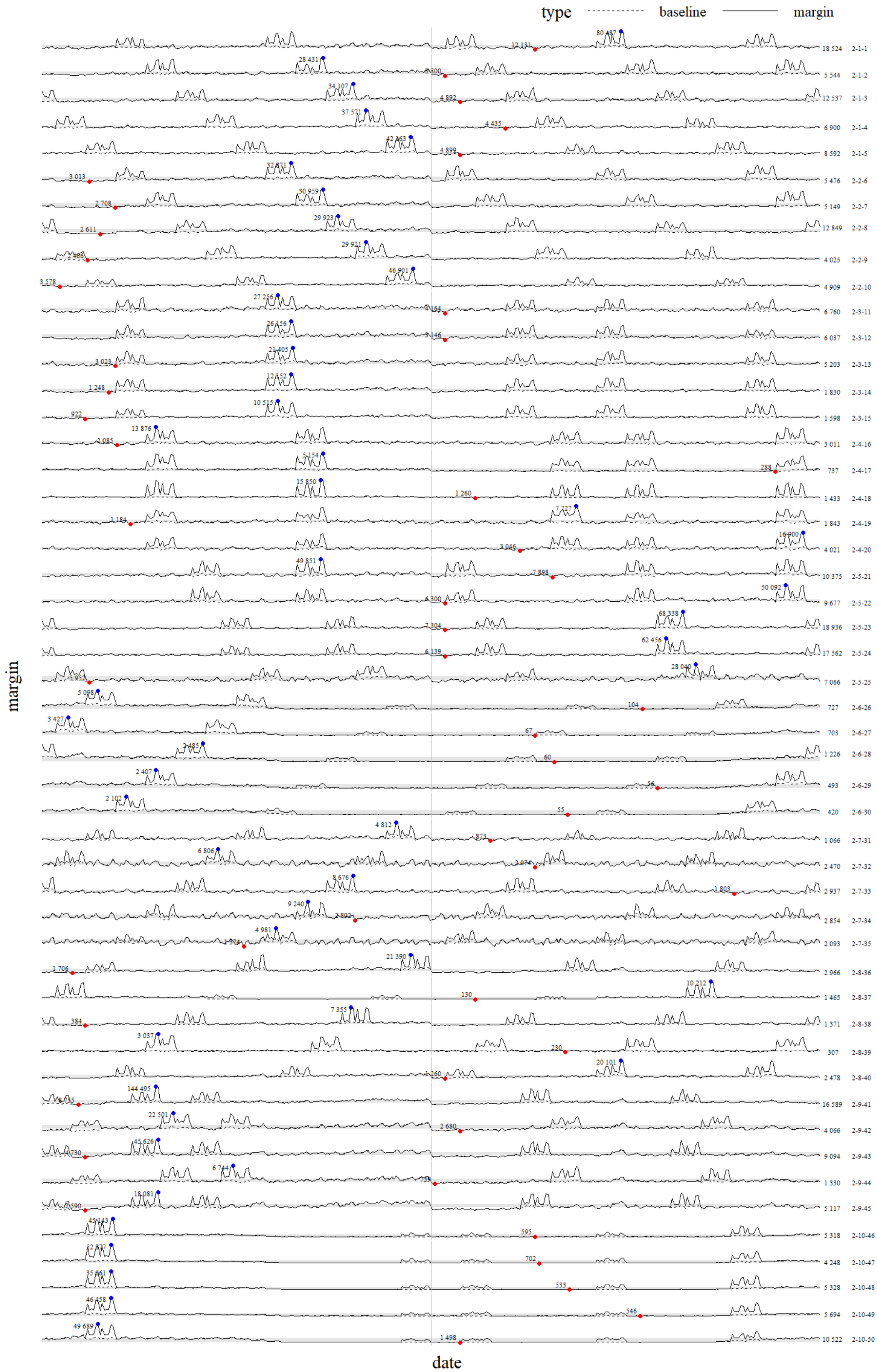


Figure 1.7. Region 2 – Daily product sparklines of the base data set over one year.